ABSTRACT
                The first part of this dissertation is a
metatheoretical discussion of the needs and means of semantic
analysis. This discussion includes sections on metalanguage, deep and
surface structure and structural semantics, and procedures for
relating the English of abstracts to a classification concordance in
the same language. The second part describes the experiment, in which
a mechanized indexing system was developed and applied to 50
abstracts from a bibliography on documentation. In this system, each
text word or symbol is replaced by a code, after which the machine
performs step-by-step concatenations, rewriting two or more codes as
a single code, until each abstract is rewritten as a single code.
Part three is a discussion of problems of semantic representation in:
(1) theoretical background--constituent-structure rules, lexical
entries, and conjunction, discourse and relative clauses; (2)
expressions such as "a discussion of..." and "this paper
discusses..." (3) conjunction, relative changes and presupposition;
(4) asymmetric conjunction--proposes an analysis in terms of case
theory; and (5) examples drawn from the corpus that present
difficulties for current case theory (for example, "as machines
learn..."). The appendices contain descriptors, sample abstracts, and
other material used in this study. A 14-page bibliography concludes
the work. (KM)

A SEMANTIC ANALYSIS OF ABSTRACTS

Around an experiment in mechanized indexing

Mémoire présenté pour l'obtention du grade
de docteur en Philosophie et Lettres, par

Jacques NOEL
Licencié en Philosophie et Lettres
(Université de Liège, 1957)
M.A. (University of Victoria, 1968)

Université de Liège
1971-1972

CONTENTS

APPENDICES

ACKNOWLEDGEMENTS

This thesis ends a fairly eventful career as a graduate
student, a career during which I had to meet and study with
several experts in linguistics and in automatic indexing
in order to keep in touch with the two fast-moving fields
which, as early as 1964, I set out to try to bring together in
my work.

Throughout my graduate studies, the help, confidence,
and encouragements of my supervisors at the University of Liège,
Professors L. Delatte, Ph. Devaux and I. Simon have been of
incalculable value. During the preparation of this thesis,
my former teacher of English, and now my head of department,
Professor Simon gave me unique opportunity to put my views
on linguistic theories to the test by offering me to teach
her students of the licence.

I have a special debt to Professor J.P. Vinay : in 1967,
at a time when what I regard as the major recent contributions
to semantic studies had not appeared, Professor Vinay accepted
to take an interest in my mechanized indexing project and to
supervise my work towards an M.A. degree in linguistics, so
that I was fortunate enough to share his unique experience
of linguistics and his pioneering views on semantic analysis.
The mechanized indexing experiment presented in Part II of
this thesis would not have been possible without his help,
that of Peter A. Darling (Director) and of Phyl Rhynas, of the
Computer Centre of the University of Victoria, AND without
the devotion and hard work of Rémi Michelot who has so aptly
and successfully developed the computer programmes for the
experiment.

M. Jean-Claude Gardin, my discussions with him, his
detailed comments on all that I wrote before this disserta-
tion, as well as the research of his co-workers and his
thinking at the head of the Laboratoire d'Automatique docu-
mentaire et linguistique (C.N.R.S., Paris and Marseilles)
and of the Direction d'Etudes Sémiologie et Documentation
(Ecole Pratique des Hautes Etudes, Marseilles) have had
a decisive influence on my work throughout my graduate studies.

My thanks are due to all these kind people. To them
I can only offer this thesis in the hope that it will in part
repay their confidence in me and in my plans to deal linguis-
tically with the problems of indexing and its mechanization.
I also wish to express my gratitude to Colette Billiard who
has done the Herculean job of typing the bulk of this disserta-
tion and who, under the constant pressure of time, did her
job cheerfully and impeccably. I have no words with which to
thank my wife Françoise. The writing of this thesis has been
as much of an ordeal for her as it has for me, for which I
beg her forgiveness.

PART I : METATHEORETICAL DISCUSSION

1.1. INTRODUCTION

> It may be that the next great advance in the study
> of language will require the forging of new intellec-
> tual tools that permit us to bring into consideration
> a variety of questions that have been cast into the
> waste-bin of 'pragmatics'.
> (Chomsky 1969a, 81).

1.1.1. Outline

The present metatheoretical discussion is concerned
with the needs and means of semantic analysis : its purposes,
including the requirements of form which it should meet;
the role, nature, and interrelations of its components,
including linguistic theory and its subdivisions; and the
role of the linguist-observer. As I shall show in this
introduction, such a discussion is necessary, for various
reasons which have to do with the kind of topic I am
dealing with : semantic analysis of English abstracts
(summaries of scientific texts) in a particular field
(scientific information, or documentation), and the mechaniza-
tion of this analysis. The question raised in the second
section of this introduction (1.1.2.) is why, according to
some well-known standards and practices in 'pure' linguistics
my topic is open to controversy in practically all
respects, even though the same kind of topic is commonly dealt
with in such areas of language studies as content analysis
(Stone et al., 1966), and documentary analysis (in the field
of documentation : e.g. Gardin and co-workers). This
thesis as a whole may be regarded as an attempt to answer
the above question by proposing an approach to semantic
analysis in which some theoretical proposals made in linguis-
tics and some metatheoretical positions underlying the docu-
mentalists' and content analysts' semantic investigations are

regarded as complementary. As will be suggested, arguing for
this mutual relevance is in itself doubly controversial in
the present state of language studies; many linguists would
deny the semantic study of texts a status in current linguis-
tic science, while the content analysts in particular are
doubtful about the relevance of linguistic theories to their
semantic analyses of texts.

The third section of this introduction (1.1.3.) discusses
the exclusive role commonly assigned to linguistic theory
in linguistics and suggests that the arguments advanced for
such a role, which are factual and not theoretical arguments,
are fallacious as far as semantics is concerned in that they
beg a crucial question to which the semanticist should address
himself, and appeal to precisely the type of empirical facts
which make up the semanticist's fundamental problems. One of
these is the epistemological problem of how 'knowledge about
language' is attained in the form of statements about language
which owe their specific meaning and communicative function
to their being part of a whole : a particular 'universe of
discourse' - for instance, some linguistic literature -
containing the particular knowledge shared by members of a
community.

This problem, as I argue in section 1.1.4., has to do
with the broad distinction made in logic and elsewhere
between object-language and metalanguage (the language
used to describe an object-language), and around various
possible interpretations of this distinction. In fact,
after Leech (1968, and 1969, 5), one can describe the
semanticist's task as that of explaining what underlies
such statements as x means y, and a particular semantic
investigation can, at least in principle, be defined in
terms of the selection which the investigator makes of the
object-language (x) and metalanguage constituents (y), and of

the various possible substitutes for _means_ in such statements,
among the many possibilities that are open to him; in
addition, I would agree with Leech's suggestion (Leech 1968)
that the relations between the linguist-observer's state-
ments should be recognized as a fundamental problem in
semantics. More generally, I shall argue that linguistic
theories can be regarded as only _one_ among various possible
types of metalanguages, and that the relations establish-
able between these and the corresponding object-languages
constitute the central empirical facts and epistemological
problem with which a semanticist should be concerned.
Classification systems and the various other kinds of
'documentary languages' set up and used by documentalists
for purposes of scientific communication constitute the
other type of metalanguage which is given a central role
in this thesis. The proposed function of such a metalanguage
in this thesis is intended to remedy another shortcoming
(closely related to the first) of the conception assigning
linguistic theory an exclusive role in semantic analysis.
In this connection, I shall argue that a semantic metatheory
needs to give recognition to limitations that are inherent
to linguistic theory, namely that it is unable to specify
the field-specific 'concepts' or "systems of knowledge
and belief" (Chomsky) with which semanticists operate.
Linguistic theory, for instance, is unable to specify the
'meanings' which documentalists will look for in the
documentation abstracts under investigation in this thesis.

A further section of this introduction (1.1.5.) serves
to argue, after Lyons, that the linguist can and should
remain neutral towards a number of philosophical and psycho-
logical controversies about the nature of meaning and about
the goals of semantic investigations of natural languages.

The last section of this introduction (1.1.6.) outlines
the metatheoretical conception put forward in the rest of the
first part of this thesis, chiefly after Gardin 1969, and
on the basis of fairly recent theoretical developments
within and around the trends of linguistics broadly referred
to as transformational-generative (TG) theory : in particular,
work by Harris, Lyons, Leech, Fillmore, and some of their
co-workers. In brief, I shall follow Gardin 1969 in consider-
ing that semantic analysis needs to be based on three main
components and that the purpose of a particular semantic
investigation is to clarify their nature, structure, and
interrelations : a metalanguage, such as a classification
system or similar conceptual organizations used by content
analysts; a theory; and a procedure, serving ideally to
make the relations between metalanguage and object-language
explicit by means of a system of rules. Metalanguage, theory,
and procedure are further discussed in the last three chapters
of the present metatheoretical discussion : 1.2., 1.3., and
1.4., respectively. The ideal underlying this three-component
semantic analysis is, briefly, a reconciliation of 'conceptual'
communication requirements put forward outside linguistics
with the 'formal' requirements of TG linguistics.

The selection of my theoretical sources requires a word
of explanation at this stage. First of all, it may seem
that such authors as Harris and Fillmore have little in
common. This may be the case as far as their theoretical
proposals are concerned : in fact, they do not quote each
other. From a metatheoretical point of view, however, I
shall attempt to show in this thesis and in section 1.1.6.
in particular, that they are relatable in terms of a basic
dichotomy which needs to be made in linguistics between
authors concerned with the communicative function of language
and those who claim not to be, and a parallel distinction
between preoccupations with distinctive statements versus

a concern with 'general' statements, where the requirements
of distinctiveness are left (at best) for a later stage of
research. In this respect, I feel that some recent work by
Harris and Fillmore, (in particular Fillmore's recent study
of 'verbs of judging' and his explicit concern with semantic
fields : Fillmore 1969b) display similar preoccupations with
distinctive statements and with the communicative function of
language. In addition, I must stress that the selection of
my theoretical sources results from my experience of the
field of linguistics during the preparation of this study
and is governed by my judgments about the compatibility of
the theoretical proposals in question with the topic of this
study and with the metatheoretical conceptions I propose,
My position .. . in this respect is in no direct way determined
by such chance factors as influences undergone during my
training in linguistics, still less by a priori decisions
of the kind exemplified below by Botha's decision (Botha
1968) to ork within the "single" framework of TG theory.
Besides, irrespective of the research with which I may
not be familiar, many studies not mentioned in this disserta-
tion have been omitted either because I have found them
irrelevant, or because I believe them to be reducible to,
or superseded by, the studies which I refer to - at least
as far as their applicability to my topic is concerned. This
is especially the case of works belonging to various tradi-
tions of language studies which are mentioned or discussed
in my M.A. thesis (Moßl 1968), or by the linguists I refer to.
In particular, I shall not attempt to bring in several
studies in the field of logic as well as many European
contributions to linguistics which are referred to in the
publications of Lyons, Leech, and Fillmore. Similarly, I
shall not refer to several works in various fields of
language studies outside linguistics proper, because I am
not acquainted with them and because a thorough discussion
of these studies (supposing it is feasible) would have
required a world-wide survey of semantic analysis methods

which is beyond the scope of the present doctoral thesis.
In particular, I shall not attempt to discuss several
studies in the fields of literary analysis, documentary
analysis and classification research, which are referred to
in books and papers such as Coyaud 1966, Coyaud and Siot-
Decauville 1967 and Gardin 1969, and in the bibliography
of abstracts under investigation in this thesis (Janaske
1962).

1.1.2. A controversial topic.

In order to discuss the various controversial aspects
of my thesis topic, it is useful to mention a few diverging
attitudes among linguists on several important matters. These
disagreements are well-known, but they need to be mentioned
here as they have direct bearing on all the main features
of my topic, and therefore on this topic as a whole.

First, there is disagreement as to the place of
computers in linguistics. While quite a number of researchers,
including 'pure' linguists, are engaged in computational
linguistics, many linguists would refuse computers a place in
their science, or do not feel the need to use them in their
work. A striking example is the difference in this respect
between Harris and his co-workers, who have been actively
concerned with computer analyses (including the implications
of "transformations" for the problems of documentary analysis
discussed in this thesis : Harris 1970), and Harris's pupil,
Chomsky. The mechanization experiment presented in Part II
of this dissertation is intended to clarify a possible role
of the computer in linguistics on the basis of proposals
made by Harris and co-workers (after Harris's "string
analysis": Harris 1965), and of Gardin's metatheoretical pro-
posals as experimented by his co-workers, from Coyaud to
Bély et al. 1970. In the experiment which I present, a small
number of the documentation abstracts under investigation in
this thesis are represented by string formulas stored before-
hand in the computer. The machine operates with this gramma-
tically organized material and gradually reduces the formulas
recognized in each text to a single notation which stands
for one or more entries of a classification system for docu-
mentation serving as a metalanguage for this field. This,
then, is broadly the basis on which Part II of this thesis
relates linguistic theory with the documentalists' (and the
content analysts') conception of semantic analysis.

I also suggest that the mutual relevance of the two areas of
language studies beyond this point may also be established
when enough knowledge is gained on certain problems of
semantic theory for semantic representation to be amenable
to mechanization. These problems are discussed in Part III
of the present dissertation. In principle, they concern
equally the object-language under investigation (the English
of abstracts on documentation), the various possible meta-
languages (such as classification systems) relatable to this
object-language, and the relations between the semantic
representations establishable for both.

Another point of disagreement among linguists appears
in their attitudes to various methodological distinctions
(between syntax and semantics, grammar and lexicon, grammar
and discourse) and in the frequent treatment of one of the
problem areas thus labelled in methodological isolation.
Prominent syntacticians, for instance, have generally left
the semantic extension of their theories for later, or to
other investigators, and, except in the last two decades or
so, linguistic semanticists have shown little interest in
syntax and have contributed little to the semantics of
grammatical units larger than those dealt with by lexicogra-
phers. One of the main objectives of this thesis is to find a
basis for going beyond such methodological distinctions, since
my first concern is with people's ability to relate abstracts
(i.e. texts usually consisting of a title and of one or
more sentences) to one or more classification entries having
the form of a lexical unit and a definition consisting
of one or more sentences.

With respect to the linguistic study of units larger
than the 'sentence' - another controversial aspect of my
thesis - Zellig Harris occupies a unique place in linguistics.

Harris is one of the few prominent linguists to have been
consistently interested in analyzing texts and discourses,
and the only prominent 'pure' linguist to have dealt with
scientific discourse. For this reason, one might think that
the present study must draw heavily on Harris, or even be
centred on an exploration of his proposals, particularly
those on discourse analysis. This is not the case, however.
As far as I can judge from the literature I know, the
currently available illustrations and programmatic applica-
tions of discourse analysis and of related methods such as
the transformational approach (in particular those presented
in Harris 1968 and 1970) do not substantiate Harris's well-
known claim that a linguist can deal with everything
'linguistically' and need not be explicit about the semantic
criteria he uses, or at least need not wonder whether some
of the components of his analyses can be rightly regarded
as linguistic, or "grammatical". I feel that the problem,
to which much of the metatheoretical discussion of this
thesis is devoted (particularly around the notion of meta-
language), concerns all varieties of the language, though
it arises most dramatically with scientific varieties. Even
if on-going research by Harris and his pupils should present
extensive (instead of programmatic) analyses demonstrating
the feasibility of reducing facts of discourse (such as equi-
valence classes in texts) to linguistic rules (such as
transformations), one may still wonder how the facts and the
corresponding rules can be established without appealing
to some kind of culture or expertise which can neither be
'given' nor 'organized' by a linguistic theory. Even if
Harris can set up 'grammatical' rules accounting for the
similarity between members of an equivalence class in a
particular text (between, say, "the study of denaturation"
and "strong solution of urea"), one cannot help feeling that
in order to be able to establish such facts he has had to
consult a biochemist, or has himself received an extensive

training in biochemistry, since this field has not been
reduced to a linguistic theory, and probably never will be
(Harris 1968, 148-153). My point here, which is a central
one in the present thesis, is basically similar to that
made most explicitly by Mounin (1965) about Dubois's attempt
to set up semantic fields in French political and social
vocabulary on the sole basis of substitutional and distri-
butional techniques (Dubois 1963), and about his "implicit"
use of "conceptual" criteria in applying such techniques :

> En fait, on décide, entre toutes les unités qui peuvent
> "former système" avec une autre à partir d'un ou plu-
> sieurs énoncés donnés, de ne retenir que celles qui sont
> sémantiquement apparentées (la bourgeoisie, le prolétariat
> l'aristocratie, le capitalisme..., voilà l'ennemi; mais
> non l'alcool, le phylloxéra, etc.). La présentation a
> l'air déterminée par des critères formels; en réalité,
> on a réintroduit une décision d'ordre conceptuel,
> implicite, dans le circuit opératoire. Ceci n'invalide
> pas la construction de tels champs sémantiques, mais
> attire l'attention sur le fait que, malgré les apparences,
> on n'a pas réussi à délimiter le champ sur des marques
> formelles, et d'après des procédures exclusivement et
> proprement linguistiques .(Mounin 1965, 135)

My concern with documentation English runs counter to
another recognizable tendency in several contemporary linguis-
tic studies, including some on which this thesis relies
heavily (e.g. Fillmore). Many contemporary linguists have
an obvious preference for drawing their examples from every-
day language and seem reluctant to study other language
varieties, such as the sublanguages dealt with by Harris
and Dubois, or the language varieties examined in the tra-
dition of semantic fields from Trier to Lyons's study of the
role of epistemological vocabulary in the semantics of
sentences from Plato (Lyons 1963). One of the consequences
of this situation is that the present study deals with
object-language terms which are often regarded or used as
primitives in the linguistic literature : for instance,
the words _information_ or _communication_. More importantly -
and the point is particularly relevant to the study of
discourse - a scientific sublanguage must be taken as it is,

in its attested manifestations, A the examples cannot be
invented by the linguist-observe in the same way as every-
day sentences presumably can be. At least initially, the
descriptions presented in this thesis cannot be neutral with
respect to analysis and synthesis, as Chomsky and others
have claimed a linguistic theory or description should be.
This study cannot without oddity be concerned (at least
directly) with language production or synthesis, and must
give primacy to analysis : the definition of a metalanguage
(given below in Cherry's formulation : Cherry 1957, 305)
as "observer's language" reflects this primacy.

While the various forms of disagreement mentioned so far
can perhaps be treated as a fact inherent to scientific
research as a whole, one may still feel that the topic of
this thesis is difficult to reconcile with the present state
of linguistics. The abstracts under investigation may be
expected to present almost all the main difficulties faced
by past and present linguists (from most specific lexicolo-
gical problems to the use of the article in English) with
no obvious possibility of resorting to the kind of method-
ological simplifications that are possible when linguists
invent their examples, or collect them after some finite
inventory of theoretical criteria. This, one may argue,
makes the topic of this thesis hopelessly complex considering
"the state of our ignorance" (Leech 1969, 3, quoting Weinreich):
for Fillmore, for instance, bachelor is "the only English
noun that has so far been given a semantic analysis", and
this linguist feels that "the world must wait another two or
three decades before it will see anything resembling a
respectable grammar of English" (Fillmore 1968b, 28-29).
Indeed, considering the state of linguistics, various authors
suggest that progress can only be achieved by the very prelim-
inary simplifications of the field of inquiry that appear
to be impracticable in this study : for example, the present

investigation is apparently difficult to reconcile with
Fillmore's recent suggestion (ibid.) that progress in lin-
guistics is best served by trying to understand "the elementary
structure of the "propositional" core of simple sentences".
To justify such simplifications, mention is occasionally
made of other sciences which are said to owe their present
degree of advancement and diversification to their pioneers'
selection of relatively simple problems such as, say, perfect
gases in chemistry. For some linguists, it appears that such
an attitude has direct bearing on the semantic study of
texts as a field of inquiry in linguistics. As the following
passage from Langendoen suggest, the argument is roughly that
a semantic study of texts can only be misguided since it
forces one to deal with just the two areas in which linguistic
theories are assumed to be most inadequate. Thus, while he
admits that some sentences and presumably some texts clearly
belong to langue, as when a linguistic community decides
that "a section of a legal document" is applicable to "a
particular situation" Langendoen proposes to

> return to the Saussurean conception of words as belonging
> to langue, but sentences and discourses as belonging
> to parole-- not because we lack the theoretical apparatus
> to deal with sentences and discourses syntactically,
> but because we lack the necessary apparatus for dealing
> with them semantically in any systematic fashion (Langen-
> doen 1967 , 107 and 101).

It should be noted, however, that an argumentation
similar to Langendoen's can also be used to decide . what
a semantic analysis should be at the word level. This appears
most typically in Leech's recent proposal to characterize
names of animal species by means of "a single contrastive
component" ("e.g. 1 SPE for 'dog', 2 SPE for 'cat', 3 SPE
for 'elephant', etc.") and to exclude various other possible
semantic features, such as 'having a trunk', 'ivory-producing',
etc., for 'elephant'. As he neatly notes, the effect of this
decision is that the theory will class, 'This elephant is a
tiger' "as a logical contradiction" while it will treat 'This
elephant has eighty legs' as "well-formed".

Leech bases his decision on the principle that

> "an investigator has reached the limits of semantics
> whenever he is faced with an indefinite number of
> criteria or partial criteria for distinguishing the
> meanings of a set of terms within the same semantic
> field", and any choice between those criteria can only
> appear to be "arbitrary"; "to have an infinitely or
> indefinitely long semantic specification would run
> counter to the whole notion of an explicitly formulated
> semantic description we have been considering".
> (Leech 1969, 87-88).

At this point, the reasons why the topic of this thesis
appears to be controversial by several standards and
practices of current linguistic research can be clarified
by referring to a principle which Leech presents as underly-
ing his semantic investigation, and which is most obviously
relevant to Langendoen's point about the semantic study of
discourses. It is the position which consists of letting
"the theory of semantics determine the extent of its subject-
matter, rather than the other way round", where "theory"
refers to "standards internal to a semantic theory" (Leech
1969, 80 and 85; my italics). I do not mean to say that
all is clear about the relations between Leech's principle
and the various standards and attitudes of current linguists
which make my thesis topic appear as a controversial one.
What I wish to suggest, however, is that the kind of prin-
ciple which Leech formulates is the clearest thinkable way
of pinning down the reasons why my topic can be judged in
this fashion. Besides, this interpretation finds support
in a recent appraisal of the situation of current linguistics
by Botha, to be fully discussed in the next section (1.1.3.).
Finally - and most importantly, as far as the argumentation
of this thesis is concerned - the principles which Leech
and Botha formulate hinge on an epistemological conception
of the role of "theories" in language studies, which I wish
to clarify in this thesis and which happens to be similar
to the conception of semantic analysis of texts adopted
by content analysts, as I wish to show in the rest of this

section. The crucial difference, however, between the two
attitudes is that Langendoen and others would object to the
semantic study of texts because of the deficiencies of
linguistic theory while the content analysts base their
semantic analyses of texts on non-linguistic theories.

Semantic analysis of texts can be regarded as the
central pursuit of natural language investigations in the
fields of documentation and content analysis. The sociologists
and psychologists involved in content analysis and the docu-
mentalists, whose responsibility is to serve as a link between
writers and readers of scientific texts, are both concerned
with how to characterize the meanings of texts. In both fields,
essentially two approaches to this problem have been explored.
One is the statistical approach, which is associated with such
classical   studies as Berelson (1952) in the field of content
analysis, and with the works of such authors as Baxendale,
Luhn, and others in the field of documentation. I shall not be
concerned with the statistical approach in this thesis. A com-
parison between this approach and the other is beyond the scope
of this study, as several publications by Gardin, Salton, and
many others suggest, together with the many arguments advanced
for and against the use of statistical methods in linguistics.
Here, I shall be concerned only with the other approach.
However one may wish to call it, whether it is practised
by the content analysts working around Stone or by documenta-
lists such as those of the Gardin team, it can be regarded as
fundamentally the same approach; the similarity, which has
been shown in Gardin 1969, will become obvious in the course
of the present metatheoretical discussion. For the sake of
presentation, I shall devote the rest of this section to the
main features of Stone and his co-workers' approach to semantic
analysis (as discussed in Stone et al., 1966). Afterwards, I
shall concentrate on the documentalists' point of view.

The content analysts' preoccupations with recorded
speech and writing that is relevant to their psychological
and social sciences can be described as follows: "Much as the
archaeologist infers the life of a culture from the pattern of
remnants, so the content analyst infers the orientation and
concerns of a speaker, subculture, or culture from the record
of what is said" (ibid., 19). Thus, inference-making, or
drawing conclusions from texts is the ultimate objective
of content analysis. The texts which they analyze cover a
considerable variety of subjects : political (e.g. presiden-
tial speeches), psychological (e.g. suicide notes), literary
(e.g. Mark Twain's Huckleberry Finn), etc. The conclusions or
inferences are made on the basis of what is called "content
characteristics" or "categories". For instance, the content
analyst will "rate a series of editorials on whether they
are for, against, or neutral toward a particular policy".
Or, they may want to identify the extent to which "conserva-
tism", and various "characteristics" considered to be "signs
of conservatism", are expressed in texts (ibid., 4 and 7).

As in documentary analysis, the emphasis is on the need
for semantic analysis to meet certain requirements of form.
These are met by aiming at "procedures for transforming intui-
tive judgments into explicit rules" (ibid., 5), that is, rules
and procedures "that can be replicated exactly by other
analysts" (ibid., 11). Computerization is regarded as
essential to achieve this aim, next to other, practical,
objectives such as "to ensure reliability and speed of coding
while reducing its tedium" (ibid., 5); in fact, Stone et al.
1966 is essentially concerned with presenting a computer
approach to content analysis called "The General Inquirer":

> The nature of a computer program requires that both the
> categories and the rules for identifying and recording
> characteristics occurring in the data be explicitly
> stated. The computer then systematically applies the
> categories and rules to the data in a completely
> objective manner. Whether the (...) procedure was
> reasonable or best suited to the inferences being made
> may be debated, but the procedures themselves are explicit
> and clear.(ibid., 12)

To help them make their procedures explicit, the content
analysts obviously need "an adequate theory of the relation-
ship between social scientific variables" (such as "conserva-
tism" and its constituent characteristics) "and how they are
expressed in language". From the outset, however, Stone
et al. echo the linguists' pronouncements on the state of
our ignorance; they point out that no such theory is currently
available and stress the need for the investigator to "fall
back on" his "intuitive" ability as a speaker of the language
to specify as best he can "the alternative ways in which a
characteristic can be expressed" (ibid., 11 and 8) :

> we lack an adequate theory of language to direct us in
> finding the alternative signs that express a particular
> concept. In a situation where something is to be said,
> there is no theory to tell us what words will be used
> to say it. (...) There is as yet no good theory of
> symbolic communication by which to predict how given
> values, attitudes or ideologies will be expressed in
> manifest symbols (...) There is almost no theory of
> language which predicts the specific words one will
> emit in the course of expressing the contents of his
> thoughts" (quoted from Lasswell et al. 1952, 49).
> To be useful, such a theory will have to combine
> specific knowledge of the individual speaker and the
> perceived social situation, together with a general
> knowledge of language. (ibid., 9-10)

Moreover, when Stone and his colleagues refer specifical-
ly to linguistics as a possible foundation for their analyses,
they repeatedly emphasize the differences between the
linguists' objectives and their own. They do make use, it
is true, of syntactic codes and rules and contemplate a
mechanization of the "hand syntax coding" they report

(ibid., 161). But, giving the relatedness of "sit" and
"chair" as an example of the difference between "distribu-
tional structure" and "the groupings of words used in content
analysis", they stress that the similarities between their
work and that of distributional linguists such as Harris
"may be more apparent than real" (ibid., 60). With reference
to linguistic semantics, they arrive at a similar conclusion.
While they note "a remarkable operational parallel" with
Katz and Fodor's (1963) proposals, they insist that their
goals differ from that of these authors in that the purpose
of content analysis is to infer the "origin, uses, and effects
of signs within the behavior in which they occur" (ibid., 18).

Stone et al.'s attitude to linguistic theories is
obviously negative. But the content analysts' research has,
next to the requirements of form mentioned above, a positive
aspect to which I now wish to turn, and which is central to
my discussion. It is the consideration that, despite the
handicap of having no adequate linguistic theory of "symbolic
communication", semantic analysis can proceed perfectly well,
on the basis of the central role which the content analyst
gives to his social and psychological theories : "the theory
being investigated (...) determines the texts to be compared
(that is, the research design), the categories and rules
for application that must be constructed, and the kinds of
inferences that might be drawn" from the relations established
between the texts and the categories (ibid., 14-16). To
return to the example of "conservative" texts, the social
scientists' theories will serve to determine all the essen-
tial components of their semantic analyses and their extensions
except the specification of "the alternative ways in which
a characteristic can be expressed" in language : the various
constituent concepts of "conservatism", the selection of
texts that will throw light on these concepts, the relations
to be established between the texts and the concepts, and the
ultimate exploitation of the analysis. Moreover, the first
two steps of the analysis - the selection of texts and the

definition of the "characteristics" - are taken on the basis
of the theory, prior to the analysis itself.

Let us note, first of all, that this is not a claim which
Stone et al. make about the nature of semantic analysis : it
is a description of how their analyses are actually set up.
Besides, the facts revealed by Stone and his colleagues are
in contradiction with the claim made by some linguists and
mentioned above in connection with Harris that linguistic
data, and in particular texts, can be analyzed on the sole
basis of a linguistic theory. Lastly, and most importantly as
far as my argument is concerned, the linguists and the
content analysts concur in assigning theoretical considera-
tions a central epistemological role in semantic analysis.
On the other hand, there are obvious differences between the
two kinds of conceptions. Leech, say, does not insist as do
the content analysts on the need for the relations between his
'categories' and the data to be amenable to testing with a
computer. Moreover, the content analysts feel that they can
carry out their semantic analyses on the basis of their non-
linguistic theories, and do not regard the absence of an
adequate linguistic theory as a fundamental obstacle, but
merely as a handicap, while Leech assigns linguistic theory
an exclusive place, even at the cost of excluding some facts
from the scope of his investigation (as when 'This elephant
has eighty legs' is treated as well-formed by the theory).
In the following section, I shall attempt to clarify and
discuss the kind of position advocated by Leech, after a
recent study by Botha (1968).

1.1.3. Linguistic theory in linguistics.

In the opening section of his recently published disser-
tation, Botha 1968 presents an analysis of the current scene
in linguistics which, judging from my experience of the
field, I believe to be fundamentally correct. Moreover,
I do not know of any other author who has made the same
points so explicitly, a nd I suggest that they clarify Leech's
proposal to "let the theory of semantics" (and "standards
internal" to it) "determine the extent of its subject-matter,
rather than the other way round" (Leech 1969, 60). Botha, then,
points out that

> the linguistic research of an increasing number of
> present-day linguists is governed by (...) the general
> methodological or epistemological principle that
> extensive knowledge about "language" or "linguistic
> structure" can only be sought in the framework of a
> particular well-defined linguistic theory or model. I
> am not familiar with recent attempts that successfully
> prove the untenability of this principle. The case is
> rather that the linguistic argumentation of the scholars
> who do not hold this view is characterized by obvious
> methodological flaws...
> (ibid., 13; in a passage to which I shall return, Botha
> criticizes the use made by some linguists of terms such
> as 'the sentence', or 'the word' outside a particular
> theoretical framework : ibid., 81-83).

Botha immediately accepts this principle as a basis for his
own investigation, together with another principle, "most
convincingly formulated by Popper 1963", "that empirical
science can only advance when scientists attempt constantly
to refute their theories by applying them to crucial cases"
(ibid., 15). According to the above principle or principles,
the crucial choice which an investigator has to make is to
select one theory among those available. Botha does this
from the outset, by adhering to "the linguistic theory of
Chomsky and his associates", because of its unique place in
current linguistics and because of the "unusual degree of
sophistication" which the discussion of "many linguistic
questions" attains in this theory : in addition to the

epistemological principle just referred to, the "second general
view influencing the scientific activities of many modern
linguists, in one way or another, is that in the past ten
years the scene of international linguistics has been dominated
by one particular linguistic theory : transformational
generative grammar" (ibid., 13-14). While I would accept this
"second general view" as a fact, Botha's consideration of
TG theory as "a particular well-defined linguistic theory"
is more questionable, even if allowance is made for Popper's
principle. While Botha may have been correct in holding this
view at the time he was preparing his dissertation, it seems
that one could rather speak to-day of a deep crisis, if not
a disintegration, of TG theory. Rather obvious evidence
for my appraisal is provided in Bach and Harms's introduction
to the proceedings of a symposium on TG theory, which was
held at Austin in 1967 and which, it seems, was attended
by the chief representatives of the younger generation of
TG linguists :

> Toward the end of the conference, when it had become
> apparent that the general agreement did not encompass
> any currently explicitly formulated model, the question
> was raised : what should we be teaching our students ?
> Langendoen's answer seems to us most apt : we should
> give them the ability to recognize an interesting
> linguistic problem when they see one, that is, one
> which throws some light-negative or positive-on our
> conceptions of what languages in general are like .
> (Bach and Harms 1968, vi)

I do not wish, however, to quarrel about this but
would rather raise two more fundamental questions by asking,
first, how a linguist can know that he has to do with "a
particular well-defined linguistic theory or model", and,
secondly, what arguments can be advanced to uphold Botha's
epistemological principle concerning the essential role of
such a theory or model in linguistic investigation. I suggest
the first question is best reduced to the second, and wish
to turn to Botha for an answer, as I am not acquainted with
better or other arguments than those he presents in his thesis.

In fairness to the author, I must quote him at length. As
already noted, his argumentation is directed against the
loose usage which some linguists make of terms like 'the
sentence', and 'the word' of which Botha argues that they
should only be used in the technical sense which they have
in one particular linguistic theory. I quote Botha's passage
in full, except for two titles (successively, 'Theory and
Concept formation in Empirical Science', and '"The Word"
and the Transformational Theory'), and for the footnotes,
which merely provide a few additional references to the
literature of epistemology on which Botha draws :

the vocabulary of everyday speech does permit the
formulation of certain generaliSatiohs, e.g., that
wood floats on water, but that metal sinks in it,
but these generaliSations have important shortcomings.
Firstly, their constituent terms will more often than
not lack precision and uhiformity of usage, causing
their meaning to be imprecise and unclear. Secondly,
many of these statements are of very limited scope,
since they lack predictive power. Thirdly, these
generalisations, e.g., 'All crows are black', usually
either lack any explanatory power or have extremely
weak explanatory power. Fourthly, heuristic power, i.e.,
the suggestion of further avenues of research, is not
usually one of the attributes of generalisations presented
in every-day terms and formulated in terms of non-technical
concepts. To overcome these deficiencies and to attain
its aims of description, prediction and explanation,
empirical science has evolved, in its different branches,
comprehensive and highly organised systems of special
concepts, referred to by technical terms. Such systems
are usually called "theories" and the concepts in terms
of which they operate "theoretical concepts".
   The introduction of theoretical concepts is subject
to two fundamental conditions. Firstly, they must have
EMPIRICAL IMPORT (cf. Hempel 1952, p. 39), i.e. it must
be possible to relate them indirectly and in a non-arbi-
trary way to the data they "account for", or "help to
account for". Secondly, theoretical concepts must possess
THEORETICAL or SYSTEMATIC IMPORT , i.e., they must permit
the formulation of general explanatory and predictive
principles. The notions 'theory', 'theoretical' or
'systematic import', and 'empirical impor̃' can be
further elucidated by the following metaphor of Hempel
(1952, p. 36). He compares a scientific theory to a
complex spatial network : "Its terms are represented
by the knots, while the threads connecting the latter
correspond, in part, to the definitions and, in part,

to the fundamental and derivative hypotheses included
in the theory. The whole system floats, as it were,
above the plain of observation and is anchored to it by
rules of interpretation. These might be viewed as strings
which are not part of the network but link certain points
of the latter with specific places in the plain of obser-
vation. By virtue of those interpretive connections, the
network can function as a scientific theory : from
certain observational data, we may ascend, via an inter-
pretive string, to some point in the theoretical network,
thence proceed, via definitions and hypotheses, to other
points, from which another interpretive string permits
a descent to the plane of observation." The links exist-
ing between a theoretical concept and the (observational)
data constitute its empirical import, while the "threads"
linking it to the other theoretical concepts symbolise
its theoretical or systematic import.

From this exposition, and especially from the metaphor,
it follows that the content or interpretation of a
particular theoretical concept C in a theory T is only
partially, and in most cases indirectly, determined by the
data to which it is related. C's interpretation or
content can only be fully determined by taking into
account also its relationships to the other theoretical
concepts contained in T. As Braithwaite (1962, pp. 230-
231) puts it "an understanding of a theoretical concept
in a scientific theory is an understanding of the role
which the theoretical term presenting it plays in the
calculus expressing the theory...". Furthermore, from
this view follows the corollary that theoretical concepts
"cannot be understood apart from the particular theory
that implicitly defines them" (cf. Nagel 1961, p. 87).
Let us now examine the cited criticism of Uhlenbeck,
as well as his remedial suggestions, against this method-
ological background. Note that Uhlenbeck formulates
his argument in a way to which the following expressions
are crucial : "THE linguistic sign", "THE two basic
units of language", "THE problem of the relationship
between syntax and semantics", "THE unique position of
THE word as a linguistic unit", "THE sentence", "sentence
and WORD are correlative units". From this usage of
terms, note especially the use of THE , and from the
general trend of his argument it may be concluded that
Uhlenbeck holds the opinion that in a certain sense
concepts such as 'the linguistic sign', 'syntax',
'semantics', 'the word', 'the sentence' have an inter-
pretation independent of any specific linguistic theory -
he regards the insight that "sentence and word are
correlative units" as "fundamental for any sound theory
of language". That this view is untenable, follows from
the generally accepted methodological principle that
fruitful theoretical concepts can only have an interpre-
tation within the framework of a particular theory.

Uhlenbeck not only loses sight of the principle that the
empirical links existing between a theoretical concept
and the data are mostly only indirect, determined there-
fore per theory, but he also disregards the principle
of systematic import. If it were true that fruitful
scientific concepts existed independently of and prior
to particular theories the laborious efforts of scientists
to construct intricate theories would be absurd and a
waste of time and energy.

An alternative interpretation of Uhlenbeck's claim is
that the concept of "the word" and "the sentence" as well
as the relationship existing between them as formulated
by Reichling in 1935 "is fundamental for any sound theory
of language". This view is equally unacceptable when
evaluated within the methodological framework outlined
in the preceding paragraph. The so-called concepts "the
word" and "the sentence" possess empirical and theoretical
import only within the framework of Reichling's theory.
One can interpret them only when they are regarded as
theoretical constructs in terms of which this theory
operates. It should be obvious that their interpretation
is necessarily partly determined by the relationship
existing between them and the other theoretical concepts
in Reichling's theory, concepts such as 'taalsysteem',
'taalgebruik', 'uiteenplaatsbaarheid', 'volstrckte
vormelijke bopaaldheid', 'semantisch aspect', etc.
(cf. Reichling 1965, pp. 24-58). Thus, on account of
their systematic import it is already impossible to use
Reichling's concepts 'word' and 'sentence' in the trans-
formational theory, or in any linguistic theory differ-
ing significantly from Reichling's theory. Note that even
if the word concept used by transformationalists
were to be formalised, it would not possess empirical
and systematic import similar to that of Reichling's
concept ".the word".

Uhlenbeck has not come to grips with the idea that
the problem concerning the relationship between syntax
and semantics as discussed by Chomsky is only relevant
to the transformational theory and that it can only
be "solved" in the framework of this theory. Neither
does he appreciate the fact that there exists no obvious
basis for comparing theories of language whose aims are
as disparate as those of his own and Chomsky's.
(Botha 1968, 81-84).

I cannot help thinking that the whole of Botha's
discussion on theories really amounts to a handful of well-
known facts, which boil down to popular ideas about technical
terms and related matters, and which, as far as linguistic
theories are concerned, could have been more briefly formul-
ated (and more interestingly, as I shall try to show in the
next section) by resorting to the standard distinction
between object-language and metalanguage, i.e. between the
"language under observation and study", and "the language
used by an observer for describing an observed object-language"
(Cherry 1957, 305). What Botha says, then, can be put briefly
as follows. Theories are set up to overcome the inadequacies
of everyday language as a metalanguage. A theory is a meta-
language whose constituent concepts can only be understood
in relation to the other concepts of the same theory, which
"implicitly" defines its concepts. Besides, the interpreta-
tion of theoretical concepts is determined by the theory
itself, and "only partially, and in most cases indirectly"
by the data, i.e. the object-language in the case of a
linguistic theory. Consequently, metalinguistic concepts
belonging to different theories, and each theory as a whole,
are not reducible to one another. They can at best be compared
when they have similar "aims".

In spite of his contention that he can demonstrate
"obvious methodological flaws" (ibid., 13) in the argumen-
tation of the scholars he criticizes, what Botha provides
is not an argument, as the chief point he makes rests on
a metaphor. There is nothing objectionable about analogies
as a means of intellectual or other perception, on the con-
trary. But there is something ironical about a linguist
like Botha (who adheres to a linguistic school with well-
known requirements of form) having to present a metaphor
as an argument. In fact, the irony does not stop there.
Botha adheres to the TG theory of his time, and in particular

to the semantic component of it as set up by Katz and his
associates ; according to this theory, the meaning of lexical
items is defined by means of a "universal" (i.e. language-
independent) "theoretical vocabulary" of semantic "markers"
(such as 'human', 'male', 'adult' and 'never-married', for
the word bachelor : Katz and Postal 1964, 14) which are
posited to be "a language-independent means of representing
the common conceptual system underlying communication in
natural languages" and to be "part of the cognitive structure
of the human mind" (Katz 1967, 129). I cannot help detecting
a contradiction here between Botha's theoretical and epistemo-
logical positions : between his acceptance of Katz's
hypothesis about the universal validity of theoretical
constructs like the semantic markers, and his insistence
on the existence of sharply distinct theories within the single
discipline of linguistics and on their being largely irredu-
cible to one another. But the most significantly ironical
aspect of Botha's discussion as far as this thesis is
concerned is that his attempt to defend the use of TG theory
as a single well-defined linguistic theory forces him to
appeal to the very facts which this theory (and, for that
matter, all the other linguistic theories) have left
largely undiscussed : conceptual problems, problems of
discourse structures, and their interrelations. Typically,
for Botha a theory is something that 'hangs in the air' ;
he nowhere speaks of theories in their linguistic manifesta-
tions as texts or discourses, and his recognition of the
absence of reliable scientific knowledge on conceptual
facts within TG theory is only indirect : "even if the word
concept used by transformationalists were to be formalised".
What one would at least have expected of Botha is explicit
recognition that the metatheoretical position which parallels
the well-structured theory of Chomsky 1965 to which he
adheres (particularly its division into a syntactic, a semantic,
and a phonological component) is basically indeterminate -
as Chomsky himself most explicitly states :

"In general, one should not expect to be able to delimit
a large and complex domain before it has been thoroughly
explored. A decision as to the boundary separating syntax
and semantics (if there is one) is not a perequisite
for theoretical and descriptive study of syntactic and
semantic rules. On the contrary, the problem of delimita-
tion will clearly remain open until these fields are
much better understood than they are today. Exactly
the same can be said about the boundary separating
semantic systems from systems of knowledge and belief.
That these seem to interpenetrate in obscure ways has
long been noted. One can hardly achieve significant
understanding of this matter in advance of a deep
analysis of systems of semantic rules, on the one hand,
and systems of belief, on the other"(Chomsky 1965, 159).
More recently, Chomsky has noted that "it is not clear
at all that it is possible to distinguish sharply between
the contribution of grammar [i.e. linguistic theory] to
the determination of meaning, and the contribution of
so-called 'pragmatic considerations', questions of
fact and belief and context of utterance" (my italics :
Chomsky 1969a, 80). In connection with this opposition
between 'pragmatics' and the study of abstract formal
structures and relations, he further notes : "It may
be that the next great advance in the study of language
will require the forging of new intellectual tools
that permit us to bring into consideration a variety
of questions that have been cast into the waste-bin
of 'pragmatics'" (ibid., 81).

These passages from Chomsky need to be quoted in
connection with Botha's discussion since the question which
they raise of what linguistic theory is about cannot be
separated from the one raised by Botha about the role of a
linguistic theory in achieving 'knowledge' about 'language'.
Chomsky's remarks, as I read them, carry important suggestions
in this connection. While leaving the question largely open
for discussion they suggest that it is possible to assign
a specific role to linguistic science, that linguistics
cannot be equated with a 'general epistemology' dealing
with all (aspects of) "systems of knowledge and belief"
including itself, and that "a deep analysis of systems of
belief" can throw light on the other issues.

Moreover, Chomsky's conception of the indeterminacy
of current linguistic metatheory has a dimension beyond the
suggestion that 'conceptual' systems ("systems of knowledge
and belief") are relatable in some poorly understood way
to the facts accounted for by the syntactic and semantic
components of his theory. If all 'conceptual' systems
display the same diversity and irreducible individuality
as those manifested in the theoretical disagreements or
differences of outlook among scientists, and linguists
in particular (and there is no reason why this should not
be the case), the question arises of how general a linguistic
theory can be that incorporates conceptual facts other than
its theoretical concepts, either by virtue of some form of
metatheoretical indeterminacy, or as a result of some
consciously made decision to incorporate them in linguistic
descriptions. I shall return to these important questions
in the course of this study, by providing an example of what
can be done within a metatheoretical conception which is
related to the way I have interpreted the above quotations
from Chomsky.

Lastly, since Botha seems most knowledgeable about
current work in logic, he might also have shown that the
logicians are as helpless as the linguists about the problems
of discourse and the related 'conceptual' problems, as
Linsky suggests in the following passage about various types
of discourses, and could also have done about scientific
literatures and theories :

> In speaking about movies, plays, novels, dreams,
> legends, super titions, make-believe, etc., our words
> may be thought of as occurring within the scope of
> special 'operators'. Let me explain. Watching the
> western, I say, "I thought the sheriff would hang
> the hero, but he didn't". The context in which these
> words are said makes it clear that they are occurring
> under the 'in-the-movie' operator (...) Similarly,
> I might say, "Leopold Bloom lived in Dublin'.
> Is this true or false ? Obviously it depends upon

whether my words are or are not, within the scope
of the 'in-the-novel' operator (...) I am unable to
explain what is at issue except in terms of the
'pictures' I have discussed. (Linsky 1967, 126-7).

I have quoted Botha, and carried out the discussion of
his positions and justifications for them at somewhat tedious
length because I wish to suggest that they point to an
important generalization which Botha misses. The point I
wish to make is this. To me, Botha and everybody else in
linguistics or elsewhere, has failed to find real arguments
that would justify the linguist's scientific activity as it
is commonly pursued (and in particular the common practice
of applying a single theory to one's data), and this failure
suggests the need to recognize en alternative to Botha's
"epistemological principle". Such an alternative would
involve decisions more fundamental than those involved in
Botha's and related positions, while it would seek to provide
arguments justifying what seems intuitively sound about Botha's
principle about the use of a (possibly single linguistic)
theory in language analysis. The existence of such an
alternative is suggested by the facts discussed in the pre-
ceding section concerning the content analysts' use of their
social theories as a basis for setting up their semantic
analyses, together with their indirect recognition of the
need for a linguistic theory that would be suitable for their
purpose. Such an alternative is discussed in the following
section (1.1.4.), around the suggestion that all theories
that are applied to natural language constitute metalanguages,
and that the fundamental problem for the semanticist (and
possibly for others) is to systematically relate object-
languages to the relevant metalanguages, as well as to
relate the constituent concepts of these metalanguages
to each other explicitly, rather than be satisfied with
the recognition of "implicit" relations between them, as
Botha after Nagel apparently is.

Besides, this thesis will propose to give recognition
to both the facts brought to light by the content analysts
and others (the use of at least one non-linguistic theory
in semantic analysis) and those discussed by Botha (the use
of linguistic theory in linguistics), thus raising the problem
of determining the respective roles of the two kinds of
metalanguage in semantic analysis : the linguistic theory
or theories or parts of them on the one hand, and the non-
linguistic theories applied to linguistic data or the meta-
languages drawn from these theories on the other. I shall
return to this and to other problems in the outline of my
metatheoretical positions presented in section 1.1.6. Here,
I would merely point out that the concrete form in which
the above problem will be raised in this thesis will consist
of asking what are the respective roles of linguistic
theory and of a kind of classification in a semantic analysis
of English abstracts in the field of documentation.

1.1.4. Types of metalanguages, nd problems of communication.

A metalanguage can be defined broadly as a language
of "signs referring to signs", of "signs of signs", or
perhaps more aptly as a language of symbols referring to
symbols, if we wish to stress the "conventional" nature of
metalanguage and object-language signs as well as of the
"coordination" between the two (Reichenbach 1947, 9 and 4).
Reichenbach illustrates the notion of metalanguage as follows:
"Thus the word 'word' refers to signs ; so do the words
'sentence', 'clause', 'phrase', 'name'. .c say that signs
of signs constitute a language of a higher level, which we
call metalanguage ; the ordinary language then is called
object language" (ibid., 9). He also stres es the metalin-
guistic nature of definitions, by giving various paraphrases
of a definition of 'submarine' : for instance, "the term
'submarine' is to have the same meaning as the term 'ship
capable of going under water'" (ibid., 20). The interest
of Reichenbach's general definition and illustrations for
the present purpose is that they suggest the existence of
various kinds of metalanguage. Some may occur in everyday
language, by virtue of the general definition as "signs of
signs" and also of the obvious, but poorly understood, fact
that "the metalanguage is in the language", as when an
English-speaking linguist uses English to describe English
(see Harris 1968). Thus,next to the linguist's "observational
statements" (Leech 1968), a host of everyday expressions
such as various forms of paraphrases may be called at least
partly metalinguistic : "informal judgments" such as "It
was the same old story" (Stone et al. 1966, 4), or phrases
such as "i.e. ...", "to put it more (+ some adverb) ...",
etc. As Reichenbach notes, terms typically derived from
grammatical theories such as 'sentence' and 'clause' are
also metalinguistic, as are typically semantic statements
such as definitions.

From what precedes, it follows that any symbolic system
used to speak about natural language is by definition a meta-
language. This, then, is the case of linguistic theory, as
well as of the systems of "categories" or "characteristics"
which content analysts draw from their social theories to
carry out their semantic analyses of texts.
This is also the case of the various kinds of
documentary languages set up and used by documentalists to
represent the content of scientific documents, the best
known form of these documentary languages being classifica-
tion systems with words as entries. Perhaps the metalinguistic
character of documentary languages and their entries appears
most strikingly in these crudest forms of documentary
analysis, as the differences between the object-language
and the metalanguage conventions   are then most apparent :
when, for instance, one or a few words drawn from a classi-
fication or subject-heading list serve to characterize the
meaning of a whole book.

What has been said so far raises two main questions.
One is whether the distinction between object-language and
metalanguage is necessary and how it can be justified. The
other is whether various types of metalanguages can be dis-
tinguished, if the rather heterogeneous facts exemplified
above are not to be lumped together under the general
category of metalanguage. These questions, as well as
Reichenbach's definition, can be clarified by referring to
Cherry (1957, 305). Cherry defines a metalanguage as
"observer's language" or as "the language used by an observer
for describing an observed object-language". Moreover, and
most importantly, he relates his distinction to a communication
process by distinguishing between an "external observer"
who is "quite detached from the communication event he is
observing" and a "participant observer" who "reports upon
communication events in which he is one partner".

The first question, on how to justify the distinction between object-language and metalanguage, can be answered on the basis of Cherry's definition. It is simply a _fact_ that people can act as object-language user's _and_ as language observers, and that their symbolic conventions in the latter role differ from those in the former. Until the difference can be specified scientifically, in linguistic science or otherwise, the distinction under discussion must rest on observations such as Botha's on the existence of, and differences between, linguistic theories, and between these and everyday language, or other observations such as: Stone and his team have set up and used the _Harvard III Psychosociological Dictionary_ and others (listed in Stone _et al_. 1966, 140) as metalanguages for content analysis ; Lawlor and his team have set up a list of 68 definitions and draw from this metalanguage for documentary analysis to characterize "involuntary confession cases" in American law court decisions, etc. The distinction is also based, as far as semantic analysis is concerned, on well-known "semantic anomalies" which exist in any natural language (NL) :

> a) different NL terms are held as equivalents (_synonymy_);
> b) to a single NL term are bound several meanings
> (_homonymy_, _homography_, _polysemy_, etc ., _e.g._,"circula-
> tion", traffic, blood, money, libraries, etc.); _c_)
> different syntactic turns of phrase are held as
> equivalent in regard to the underlying logical relation-
> ship (_allotaxics_, _e.g._, "actions of _x_ on _y_", "role of
> _x_ in _y_", "_y_ is affected, influenced, etc., by _x_",etc.);
> d) a specific phrase-structure in NL may convey different
> logical relationships (_homotaxics_ ; noun-noun
> which may sometimes qualify, "lung cancer , and some-
> times be causal, "tobacco cancer", etc.) ; _e_) more
> complex equivalences are set forth between words and
> phrases of a single language (definitions, e.g.,
> "hypothermia", "a temperature lower than normal", etc.).
> (Gardin 1969, 20)

Again in Gardin's words, the recognition of such facts suffices for "the postulation of a reality of symbolic systems outside the natural languages under consideration" (_ibid_., 21), and therefore different from them – to an extent to be investigated :

To say that, in a given field of observation, x and y
will be held equivalent to each other, for example
"clock" and "timepiece", is to refer implicitly or
explicitly to a metalinguistic entity, which is made
to correspond to the x and y of the NL under consider-
ation. It makes no difference if this entity borrows
its name from that same NL, since the only important
thing is its definition (...) It often happens that a
given NL vocabulary does not provide any suitable term
for this or that entity of the metalanguage. The same
reasoning holds in inverse cases of lexical dissimilation
(a single NL term made to correspond with several
distinct entities, each designated by a different symbol
of the metalanguage) (and also of) the syntactic phenome-
na noted above, under the headings of allotaxies and
homotaxies (...) If, for instance, one establishes an
equivalence between locutions like "effect of x on y",
"the role of x in y", "y is affected by x", etc., it
implies the definition of a corresponding syntactic
unit in the metalanguage of the analysis, however one
chooses to express it. (<u>ibid.</u>, 22)

As will be noted, Gardin's points throw light on Harris's
statement that "the metalanguage is in the language". First,
the theoretically important qualification needs to be made
that the units appropriate to name a metalinguistic entity
are not always available in the object-language considered.
Second, there is no contradiction between the distinction
under discussion and Harris's observation : both 'the
language' (say, English) and the metalanguages (based on
English) are just facts. The problem is simply how they
relate to each other.

One further important aspect of metalanguages to be
stressed here,   which provides further justification for
the distinction under examination, is that metalanguages
can - to a point which research will have to establish -
be described as 'languages'. I use the term language informal-
ly to suggest the need to analyze metalanguages as <u>wholes</u>.
This point, as far as I can see, has received more illuminat-
ing attention in the field of documentation than elsewhere,
particularly with Coyaud 1966. I shall return to the problem
later, from a rather different perspective from Coyaud's,

namely the individuality or distinctive nature of meta-
languages in documentation, as it could be described on the
basis of a linguistic theory. Here, I would just say that
my metatheoretical position in this respect is similar to
that adopted by Botha in connection with the differences
between linguistic theories : the differences between
documentary languages, as I shall argue below, is a basic
fact about them which needs accounting for. Besides, pos-
sibilities of converting one documentary language in a given
field into another in the same field can be investigated,
and in fact have been, as I shall point out later. In parti-
cular in Part III of this thesis, I shall try to substantiate
the view that linguistic theory can serve to describe a docu
mentary language distinctively, i.c. as representing what is
distinctive about the corresponding 'universe of discourse'
or 'field' of the objec-language, and that the kind of
structural discussion of documentary languages given by Coyai
on the basis of a theory or model in the field of documenta-
tion (SYNTOL : Cros et al. 1964) can also be usefully
attempted on the basis of recent developments in linguistic
semantics.

To take this discussion a step further, I propose that
Cherry's definition and his reference to a communication
process can serve as a basis to distinguish various types
of metalanguages, by suggesting a relation between the meta
language used and the type of observer, as well as the need
to distinguish between a "detached" observer and one who
participates in the communication events he is observing.
Before I can attempt to distinguish various types of meta-
languages, however, one preliminary point needs to be made.
The relation I establish here, after Cherry, between the
notion of metalanguage and the process of communication
does not result from any commitment to philosophical contro
versies about the role of communication in the attainment
of 'knowledge about language'. I do not wish, either to

"dogmatically" defend this role (as Strawson admits, he does: Strawson 1950, 32) or to adhere to the opposite view (as recently expounded after Chomsky by Moravscik 1957). My position in this respect simply consists of accepting as a fact the documentalists' conception and practice of semantic analysis as an approach to communication between writers and readers of scientific texts. Such a position simply consists of giving recognition to the fact that, irrespective of philosophical controversies and scientific formulations of the process, documentalists semantically analyse texts by assigning them metalinguistic representations that are intended to be used by the users of libraries (documentation centres, etc.) to formulate questions (search requests, as they are sometimes called) which will give them access to the documents stored in the library and to the information contained in these documents. The advantage, I would argue, of giving recognition to this fact is that it is a fact that can be observed and investigated in a concrete communication situation which is natural in the sense that addressors and addressees are distinct, unlike in the 'class-room' situations commonly investigated by linguists. The fact that I have not been able, for practical reasons, to test my semantic analyses does not affect the general point I am making here. In the course of the present metatheoretical discussion, I shall present the adjustments of Chomsky's competence/performance distinction which I believe to be required by a recognition of the documentalists' approach to semantic analysis, and of the general fact of communication underlying it.

The first distinction between types of metalanguages which I wish to put forward on the basis of Cherry's definition is one between what might be called private and public metalanguages. An example of a private metalanguage is suggested by Botha's discussion : imagine, for instance,

a linguist who would use terms such as 'the sentence' in
his own way, without referring to the use of such terms in
an existing theory. Examples of public metalanguages are :
linguistic theories, the sets of "categories" used by content
analysts in their semantic analyses, and the documentary
languages such as classification systems set up by documen-
talists and used in libraries and documentation centres for
purposes of scientific communication. I shall show later
that the kind of classification system used in this thesis
meets the description of a public metalanguage in both
respects : in terms of the way it was set up, and of the
use for which it was destined. The following is evidence
for regarding the content analysts' categories as similarly
public, in the sense that they rest on agreement or consensus
between the members of the research teams that set them up :

> Suppose we want to identify the extent to which "con-
> servatism" is expressed in writing or speech.
> Investigators would undoubtedly differ in their
> intuitive interpretations of the meaning of this concept
> (...) However, by thinking about and discussing the
> different aspects of the concept, they might well agree
> on characteristics they would consider signs of
> conservatism. (Stone et al. 1966, 7).

A second distinction I wish to propose is between meta-
languages set up to account for the communicative function
of object-languages and those which appear not to be set up
for this purpose in view of the fact that they reflect a
"detached observer"'s point of view. In order to explain
this distinction, it seems necessary to posit that a consider-
ation of the communicative function of language imposes
a distinctiveness requirement on semantic analysis. This
assumption can be put differently as follows : in order to
account for the communicative function of language, a
semantic analysis must state what is distinctive about
meanings in the object-language and its manifestations.

I personally am inclined to regard the above assumption as
a truism. It needs to be stated, however, because the
facts which underlie it do not seem to be generally
recognized. The point as I wish to put it can be illustrated
as follows : in a communication situation in which the
addressor uses an object-language (say, an utterance in
chemical English) which the addressee (say, a four-year-
old child) doesn't know, the addressee will be unable to
state what is distinctive about the meaning of the utterance
in question ; he can (at best) give a vague and incorrect
semantic analysis of it (cfr. various discussions of
Harris 1968 quoted below).

The above position raises the question, which is
crucial in the present study, of how to select a meta-
language meeting the distinctiveness requirement in question;
how this metalanguage is set up, and where it can be found.
The problem may be put as follows : does there exist a
theory that would provide a suitable metalanguage, or criteria
for selecting one, and for selecting an object-language ;
if not, how can the problem be solved metatheoretically.

In order to answer these questions, I submit the third
distinction I wish to make between metalanguages : a
distinction between metalanguages set up to account for the
communicative function of language in the whole of a language
and possibly in all natural languages, and field-specific
metalanguages, which are set up and can be used directly
by a type of "participant observer" : for instance, a
physicist observing the language of physics, or an expert
of everyday life observing everyday language, as in the
case of everyday language philosophers and linguists.

As the last point is intended to suggest, I propose
that an examination of the problems under discussion is
obscured by the common practice among linguists and other
students of language to concentrate on everyday language.
The problems are obscured because in a linguistic study of
everyday language, the linguist accidentally plays two roles :
that of a linguist-observer and that of an expert who is
acquainted with what Chomsky calls "the systems of knowledge
and belief" of the field he deals with, or, more briefly,
with a given culture. The present study of a non-everyday
variety of English and the notion of field-specific meta-
language are intended to put an end to this confusion.
Besides, this notion of field-specific metalanguage has the
advantage of reducing 'cultural' facts which underlie the use
of the word field (or such logically oriented notions as uni-
verse of discourse) to a researchable form. What remains
to be seen is whether this notion is necessary, in particular
whether a linguistic theory could not provide a field-specific
metalanguage that would serve the same distinctiveness
requirements as those served by a field-specific metalanguage.
The answer proposed in this thesis is no, and the arguments
provided for it are of two kinds : first, factual arguments
about the absence of a metalanguage in linguistics for the
field under investigation (documentation), about the existence
of several metalanguages for different fields and even for
a single one (such as documentation : in this, my position
is analogous to Botha's on linguistic theories), and about
the impossibility of reducing the existing metalanguages
to ONE field-independent AND distinctive metalanguage; second,
the definition of the role to be assigned to a linguistic
theory allowing a place to the communicative function of
language.

Concretely, the semantic metatheory proposed here
after Gardin will result in using, and in determining the
interactions of two 'metalanguages' (This will be discussed
below, in section 1.1.6.). One is a 'public', and 'field-
specific' metalanguage set up for communication purposes :
a kind of classification system for documentation, the field
under investigation. The other is similarly a public meta-
language namely a form of linguistic theory of which I shall
argue that it can serve to meet the distinctiveness and
communication requirements discusses above, and can ideally
do this for a language as a whole.

As will be noted, an important consequence follows
from the metatheoretical position delineated above after
Gardin 1969 and the content analysts. The generally
acknowledged absence of any adequate linguistic theory for
semantic analysis of texts is of no consequence as far as
the availability of a 'public' and field-specific metalanguage
is concerned, and a linguistic theory cannot provide such
a metalanguage in any case. In this respect, then, the
present semantic study of abstracts in the field of documen-
tation can proceed in the same way as, say, a study in
syntax which draw the needed metalinguistic entities from
a public metalanguage (a syntactic theory). The problem in
this study is similarly one of applying and adjusting as
explicitly as possible a finite inventory of metalinguistic
categories and relations to the indefinite number of possible
texts in documentation English, on the basis of a finite
number of observations (cp. Chomsky 1965 on the competence/
performance distinction, to be discussed below).

That linguistic theory does not provide a metalanguage
such as that found in a classification system for the field
of documentation hardly needs to be shown ; what I wish
to further argue here, however, is that it is similarly

unable to provide criteria for choosing one or the best one,
for selecting the object-language, or for defining what I
would call observational happiness (or appropriatness) condi-
tions, that is for determining whether or to what extent
a given metalanguage and a given object-language are compa-
tible. In order to show this, I must turn to a point of
linguistic theory. After Austin 1962, researchers such as
Harris (1968) and Ross (1970) have posited, and investigated
the structure of, what may be called 'implicit performatives'.
Sentences such as "I sentence you to two weeks' imprisonment",
"have, instead of truth values, various conditions pertaining
to appropriatness of use" (Ross 1970, 222), as Austin had
noted - "I", for instance, needs to be some kind of judge
if the sentence is to be appropriate. By analogy with such
sentences, linguistic evidence has been found to suggest
that any declarative sentence is, structurally speaking,
a clause functioning as a complement in a sentence of the
type "I say to you ...". What I wish to argue, assuming that the
above hypothesis is correct, is that linguistic theory is
unable to specify the 'correct' identity of the "I" and of
the "you", the participants in the communication process.
The status of the you, in particular, seldom shows up in
actual linguistic utterances, as it does in some book titles,
when some appropriate specification is expressed about the
status of the reader : cp., say, "Microbiology", "Advanced
microbiology", "*Advanced microbiology for the elementary
school". As the status of the addressee, and therefore of
the observer, and of the addressor, cannot be specified by
linguistic theory, it follows that, linguistically, all
possibilities are open, including some that seem intuitively
incompatible with the communicative function of a given
utterance. To take the example of the documentation abstracts
under investigation in this thesis, a linguistic theory cannot
rule out the possibility of having as addressors and
addressees, say, four-year-old children, janitors, aged

theologians, etc : a novelist, for instance could actually do this, but what is claimed here is that the suitable meta-language for describing his novel would have little to do with a metalanguage used by documentalists.

If I am correct, the only basis for selecting the object-language and the metalanguage is the one adopted in this thesis. For the choice of the object-language, I rely on experts (Janaske and his team : Janaske 1962) who have stated that their bibliography of abstracts deals with documentation" or rather with "science information" and with various other metalinguistic entities given in the preface to their bibliography (see chapter 1.2., below). Similarly, I have chosen as a metalanguage a kind of classification system set up by its editors and their co-workers for the field of "scientific information" (N. Gardin and Lévy). My choice, then, of the object-language (or rather of its manifestations, in the abstracts of the Janaske bibliography) and of the classification used as a field-specific metalanguage rests basically on the hypothesis that in the two above judgments (Janaske's and N. Gardin's) the terms "scientific information" and "science information" are equivalent. The linguistic problem, I would argue, begins at the point where one attempts to account for the relations between such metalinguistic entities as "science /scientific information" and others corresponding to them, such as the entries of the classification system, and for the relations that are empirically establishable between these metalinguistic entities and the abstracts under investigation : when, for instance, one can say that a given abstract "deals with" a given classification entry.

Linguistic theory, then, cannot set up or select a metalanguage for the field of documentation, simply because the _establishment_ of the "systems of knowledge and belief" belonging to this field (or to any other) lies outside the _linguists'_ responsibility and competence. Similarly, linguistic theory cannot establish to what extent a metalanguage for this field is 'appropriate' as this can only be done by one or more experts of the field ; it cannot, for instance, establish the adequacy or inadequacy of the following definition of _documentation_ :

> the assembling, coding and disseminating of recorded knowledge comprehensively treated as an integral procedure utilizing semantics, psychological and mechanical aids, and techniques of reproduction including microcopy, for giving documentary information maximum accessibility and usability. (Webster's Third New International Dictionary, 1967).

Nor can linguistic theory specify which metalinguistic

entities or concepts are 'primitive' or 'basic'. As Fillmore has put it, for a linguist to decide that 'human', 'parent', 'female','born', 'married', for instance, are "semantic primitives" is to define them in terms of (his or somebody else's beliefs about) "phenomena of the real world" rather than of "other linguistic concepts" (Fillmore,1968c, 34), and I argue that a linguistic theory provides no basis for such a decision. Similarly, and again after an example from Fillmore, I suggest that linguistic theory does not set a "stopping place" at which one can be sure that one has stated 'all' the "selectional" components or features assignable to a word (say _the wolf_) in order to specify the words that can appropriately be used with it in a sentence : e.g. _ate_ in _the wolf ate_, where _wolf_ would be assigned a feature 'animal' selecting _ate_ and the like. (Fillmore 1969b, 93).

More importantly, I argue, after Bolinger's criticism
of Katz and Fodor 1953, that a distinction between "knowledge
of the language" and "knowledge of the world" (and similar
ones) is metatheoretical and that any attempt to give it a
foundation in linguistic theory is misguided and can only
be arbitrary. This I also believe to be the case of the type
of position mentioned above in connection with Leech, where
the theory acts as a filter : anything that is theoretically
untractable or incompatible with criteria internal to the
theory is 'filtered out' as lying outside the scope of
linguistics ("linguistic competence").

> To return to Katz and Fodor, it will be remembered
> (1953, 178) that they illustrate the dichotomy "know-
> ledge of the language" versus "knowledge of the world"
> by claiming that, unlike the concepts expressed by
> semantic markers such as (Male) and (Female), the
> information enabling us to understand horse shoes
> (or rather horseshoes, as Bolinger points out) as
> shoes for horses and alligator shoes as shoes made
> from alligator skin belongs to our knowledge of
> the world and must therefore be excluded from the
> theory. With Bolinger, I suggest that Katz and Fodor
> are doubly wrong : "Where do markers like (Animal),
> (Physical Object), (Young), and (Female) come from
> if not from our knowledge of the world ?" AND "What is
> strange about (Shoe-wearing) as a semantic marker-
> not as general, surely, as (Female), but general enough?"
> (Bolinger 1965, 568).

Similarly, even if a "feature" such as 'human' can be shown
to have syntactic relevance in English - for instance, to
account for who (Chomsky 1965, 150) - (but not for the
French qui), its nature, I would argue, is no more linguistic
than, say, that of 'Shoe-wearing'. In brief, I hypothesize
that all in language is conceptual, and that all the conceptual
facts are needed to explain 'linguistic structure' at
some points or other. Arguments for this view are beginning
to appear in the linguistic literature. Next to analyses of
scientific English by Harris, which are referred to in
this thesis, there are various analyses by Fillmore : he shows,
for instance, that an account of she even reads
Sanskrit requires the incorporation of some "surprise" factor

(Fillmore 1965). Let me also mention two recent papers, one by Green (1968) and one by R. Lakoff (1970) :

> Green examines English sentences with the emphatic particles either and too. She shows that for a whole range of such sentences the occurrence of these particles can only be accounted for on the basis of conceptual facts and she develops an interesting notion of implicit predication in this connection ; Green's formulation of the corresponding syntactic structures and her discussion of intonation can be omitted here. As she notes, the sentence Ho Chi Minh is responsible for a lot of deaths, and LBJ (President Johnson) is no saint either "implies that Ho Chi Minh is no saint". Similarly, Barb is seventeen, and Wendy is old enough to have a driver's license too "implies that seventeen, and therefore Barb, is old enough to have a driver's license". (Green 1968, 24). In some cases, the agreement between the two clauses can be described as "affective" : "when such sentences as He's a pacifist, and he paints well, too are acceptable, pacifist must be taken as connoting a positive attribute in the mind of the speaker if paints well is" (ibid., 20). In her conclusion, Green suggests that the syntactic component of TG grammars "will have to operate in terms of semantic representations, and will have to be sensitive to the speaker's "conception of reality" and his conception of relations between his concepts to a greater extent, and for more reasons, than has been assumed up to now" (ibid., 34). A similar point has been made more recently by Robin Lakoff (1970), in a paper which will be discussed in Part III.

Before I can turn to the question of the linguist's "neutrality" (in the next section, 1.1.5.) and outline my metatheoretical proposals (in the following section : 1.1.6.), I also wish to show that the distinctions just made after Cherry can throw light on a few other general issues. First of all, I suggest that the distinction between private and public metalanguages and the emphasis on communication can capture the facts underlying various parallel distinctions. One of them is Gardin's opposition between predetermined metalanguages and others : he speaks of metalanguages defined

"a priori", such as classifications or the content analysts'
"categories" (Gardin 1969, 26). The distinction I put
forward can also throw light on another opposition, which
may be difficult to base on observable facts. I am thinking
of a distinction between 'inductively' and 'deductively'
established metalanguages and in particular of Bach's
discussion of linguistic theories in terms of the advantages
of "hypothesis-making" or the "Keplerian" attainment of
knowledge about language, as opposed to the "Baconian"
(Bach 1965).

Secondly, the distinctions I propose can be used to
characterize the type of semantic theory made explicit by
Katz and his associates, and to refute some claims made by
these authors:

Consider the claims they make about the elementary
metalinguistic terms (the semantic markers) which they
set up. The semantic marker is a central notion since
the basic hypothesis is that the theory establishes
the meaning(s) of any (deep structural representation
of a) sentence "compositionally", that is by amalgama-
ting the semantic markers for the smallest units
(lexical items, or, better, entries) with each other,
and by deleting markers that are inappropriate (such
as 'Social Activity' for ball in The man hit the color-
ful ball), until the meanings of the larger sentence
constituents and eventually of the whole sentence are
established. What I wish to suggest first is that Katz
and his associates wrongly grant their semantic markers
a theoretical status similar to that of syntactic markers
when they claim that"a semantic marker is a theoretical
construct (...); it is on a par with such scientific
constructs as the atom, the gene, valence, and the noun
phrase. A marker like (Human) or (Color) is, then,
not an English word, but a construct represented by
one" (Katz and Fodor 1963, 208). I would argue, on
the contrary, the following : (a) When they establish
semantic markers such as (Social Activity) and (Physical
Object) for ball, they do this irrespective of the very
communication process which their theory claims to
account for (Katz and Fodor 1963, 198 on ball; Katz
1967, 129, presents the vocabulary of markers as "a
language-independent means of representing the common
conceptual system underlying communication in natural
languages"). In particular, the semantic markers

arc established by observing <u>words</u> rather than units
with which people communicate : utterances, or their
grammatical representation as sentences. In other
words, Katz and his associates are "detached observers"
in Cherry's sense. Langendoen (1967) has made a
similar point about the "detached linguist-observer"
reading ambiguities into sentences, such as 'Social
Activity' for <u>ball</u> in <u>The man hit the ball</u>. In a different
context, Lyons has stressed the <u>primacy</u> of implications
between sentences in a communication process, and the
<u>derived</u> nature of such entities as the semantic marker.
Perhaps the validity of Lyons's point and, more generally,
Katz and his associates' 'detachment' from facts of
communication are best shown negatively. Katz and
Fodor's central hypothesis is that it is possible to
set up a theory of semantic interpretation of sentences
<u>in isolation</u>, and that the particular interpretation(s)
("reading(s)") of a sentence are determined by the
"selective effect of setting" (i.e. of sentence context).
This position, it should be stressed, is <u>different</u> from
the common <u>practice</u> which consists of dealing semantical-
ly with grammatically defined material such as sentences
AS IF all about their meanings were statable in terms
of sentence-internal facts. What Katz and Fodor claim
is that such a practice can be turned into a fundamental
principle of the theory or metatheory : since "the
readings that a speaker gives a sentence in setting are
a selection from those the sentence has in isolation,
a theory of semantic interpretation is logically prior
to a theory of the selective effect of setting" (Katz
and Fodor 1963, 178). The difficulty with such a
position, I submit, is the following : one is supposed
to be able to establish the meanings of sentence
constituents (in the form of semantic markers) and
to predict "sentence readings" <u>without</u> appealing to
the very factors assumed to determine these "readings" :
sentence "setting" or context, and, beyond it, an
observable communication situation.
(b) Katz and his associates' representations of lexical
items by semantic markers can be described as drawn
from a private and 'non-field-specific' metalanguage.
This point equally applies to similar proposals made
after Katz and associates or later : e.g. Leech's
notion of "components". The only passage known to me
that throws some light on the way in which Katz and
his associates establish their semantic markers is the
following : "A semantic marker is a theoretical term
that designates a class of equivalent concepts or
ideas. Consider the idea each of us thinks of as part
of the meaning of the words 'chair', 'stone', 'man',
'building', 'planet', etc. but not part of the meaning
of such words as 'truth', 'togetherness', 'feeling',

'shadow', 'integer', 'departure', etc. - the idea that
we take to express what is common to the meaning of the
words in the former group and that we use to conceptually
distinguish them from those in the latter. Roughly, we
might characterize what is common to our individual
ideas as the notion of a spatially and temporally
contiguous material thing. The semantic marker (Physical
Object) is introduced to designate that notion. It
provides a means of expressing the generalization that
the words in the former group are semantically similar
in this conceptual respect whereas they differ in
meaning from the words in the latter group in the same
respect. Such generalizations are expressed by including
this semantic marker in the lexical readings for the
words in the former group and excluding it from the
lexical readings for those in the latter". (Katz 1967,
129-130) Rather than point out (as might reasonably
be done) on the basis of what precedes that Katz's
semantic markers are established inductively (as Bach
might say) or a posteriori (as Gardin would say),
I would argue that Katz and Fodor incorrectly grant
the same "theoretical" status to syntactic and semantic
markers and that there is a fundamental difference
between the two. A syntactic marker such as the noun-
phrase is rightly called theoretical because it is
drawn from the metalanguage of a scientific community :
the one which has set up and uses a certain type of
TG theory. The semantic markers, on the other hand,
are more rightly described as belonging to a private
metalanguage of Katz's, and his selection of the
object language (the words 'chair', etc.) as equally
idiosyncratic. This is even more obviously the case with
"the notion of a spatially and temporally contiguous
material thing", of which the marker (Physical Object)
is only a shorthand notation : the above quotation can
only be adequately described as Katz's private symbol
for his own conception of such notions as 'space',
'time', etc., and their interrelations, and in terms of
Katz's views on the applicability of this private symbol
to such object-language words as 'chair', etc. Besides,
the only communication process involved in the above
passage is Katz's appeal to the readers' consensus :
"Consider the idea each of us thinks of as part of".
In this respect, it is arguable that, unlike semantic
markers, dictionary definitions (however informal
they may be) constitute a truly public metalanguage
since people regard them as expressing the lexicological
norm or norms shared by members of the linguistic
community or communities.

The above characterization of semantic markers as part
of a private metalanguage of the authors who set them up
is intended to be more than a criticism of Katz and his
associates. The point is really this: <u>nobody, to my knowledge,
has been able to do better then Katz and associates as far as
componential analysis of words is concerned</u>. This points
to a problem which I shall attempt to clarify in this thesis,
namely the following <u>paradoxical situation</u> : while <u>non-
linguists</u> are able to set up 'public' metalanguages (classi-
fications and the like) for semantic analysis of <u>texts</u> in
a particular <u>field</u>, <u>linguists</u> have been unable to set up
more than 'private' metalanguages for semantic analysis of
other, 'smaller', types of language units such as the
sentence or the word - whether they concentrate on a 'field'
or claim to provide componential analyses valid for the
'whole of the language'. The only way out, as far as I can
see, lies in demonstrating that Lyons is correct when he
insists that componential statements are <u>derived</u> from other,
more basic, statements (such as those made in terms of the
truly <u>theoretical</u> notion of implication 'relation', discussed
below) and describes componential analysis as merely a
"technique for the economical statement of certain semantic
relations between lexical items <u>and</u> between sentences contain-
ing them" (Lyons 1968, 476). More broadly, the solution lies
in a 'structural' approach to semantics such as that advocated
by Lyons and others, as I shall argue below in section 1.3.3.

Lastly, Katz's phrases "<u>each of us</u>" and "the idea that
<u>we</u> take to..." in the above quotation embody their well-
known claim that semantic markers constitute a "universal
theoretical vocabulary" which is "part of the cognitive
structure of the human mind" (Katz 1967, 125). This claim
is best discussed in the next section, as one among other
philosophical speculations about meaning in natural language.
The linguist, as I shall argue after Lyons, should be neutral

towards these, especially when they are in contradiction with
certain facts to which he can give recognition - as is the
case for the practical impossibility of reducing all exist-
ing metalanguages for a given language or for all languages
to a single, 'universal' metalanguage. It should be noted,
however, that the claim about the linguistic irrelevance
of speculations concerning a universal semantic vocabulary
is closely tied to the empirical emphasis on the central
place of communication in linguistics, as a recent and most
extreme pronouncement by Pottier suggests : "'Pure' linguistics
is dead". The general suggestion is that linguistics, and
in particular lexicology, should turn - like sociology and
many literary studies - to the investigation of small,
"homogeneous" groups of people, because lexicologists should
concern themselves with varieties of a language "that actually
function" (Pottier 1968). Any other type of lexicological
study can merely reflect the personal views of one "detached
linguist-observer".

1.1.5. The linguist's neutrality.

According to a position most explicitly formulated
by Lyons 1968, the linguist and in particular the semanticist
can adopt a neutral attitude towards a number of philosophical
and psychological speculations or controversies about the
'nature of meaning'. As I shall argue, also after Lyons,
such an attitude goes together with a theoretical conception
of semantics discussed below under the heading 'structural
semantics' (1.3.3.). Here, I wish to deal with the first point,
and to suggest why the kind of neutrality advocated by Lyons
can and even should be adopted in the present thesis. In
keeping with the topic of this thesis, I shall draw not
only on the linguistic literature but also on some sources
in the field of documentation.

One philosophical controversy about which I shall avoid
philosophical commitment is that between 'nominalists' and
'realists' :

> Have the things to which we apply the same name, some
> common 'essential' properties by which we can identify
> them (as the 'realists' might say) or have they nothing
> in common other than the name that by convention we have
> learned to apply to them (as the nominalist might say)?
> (Lyons 1968, 401)

My position in this respect has been stated in the preceding
section : it simply consists of recognizing the existence
of certain metalinguistic symbolic conventions as a fact
worthy of research; this position, I claim, has nothing to
do with the realist/nominalist controversy as such - i.e.
as an object of century-long debates among philosophers.

Another philosophical controversy about which I shall
avoid philosophical commitment is the distinction between
"synthetic" and "analytic" statements :

> This distinction may be put as follows : a synthetic
> statement is one which is true 'contingently'
> - as a matter of empirical fact which might have been
> otherwise; an analytic statement is one that is
> 'necessarily' true, and its truth is guaranteed by (i)
> the sense of its constituent elements and (ii) the
> syntactical rules of the language. To take a standard
> example : the sentence <u>All bachelors are unmarried</u>
> might be regarded as analytic on the grounds that
> <u>bachelor</u> and <u>unmarried</u> are semantically-related in
> such a way that the truth of the sentence is guaranteed
> (...) the semantic analysis of language as it is used
> in everyday discourse need not wait upon the solution
> of the philosophical problems attaching to the distinc-
> tion between contingent and necessary truth. What
> the linguist requires is a pragmatic concep' of
> analyticity. (<u>ibid</u>., 445)

The kind of claim which Saussure makes when, for instance,
<u>la terre tourne</u> is considered as belonging to <u>langue</u>
(Saussure 1966, 173) should be qualified in the light of
Lyons's points. The same is true, I believe, of Chomsky's so-
called "mentalistic" conception of linguistic knowledge or
"competence", as an ability "intrinsic" to the speaker-hearer,
i.e. as "a mental reality" underlying linguistic behaviour
(Chomsky 1965, 4). Typically, such a distinguished epistemo-
logist as Piaget, who accepts the central place of Chomsky's
(and Harris's) <u>theory</u> in present-day linguistics, is reluctant
to follow him in his speculations about "innate" ideas and
would rather accept positions advanced by Saumjan in this
respect (on the basis of psychological and other evidence :
Piaget 1968, 77-8).

Similarly, I do not wish to commit myself to the various
other forms of competing philosophical attitudes, from those
that underlie Firth's (1964) and Bloomfield's (1933) conceptions
of meaning to those of Malinowski. Malinowski's conceptions
are particularly worth mentioning here because they embody
precisely the kind of philosophical commitments which I
wish to avoid, while they are otherwise very close to my
own metatheoretical proposals. As a recent monograph has

noted, (Langendoen 1968), there is a strange mixture of
"important true insight" and gratuitous nonsense in Mali-
nowski's presentation of his famous speculations that "even
literary and scientific language is not the expression of
thought, but its meaning is also given by correlation with the
context of situation"; Malinowski writes:

> And it seems to me that, even in the most abstract
> and theoretical aspects of human thought and verbal
> usage, the real understanding of words is always ul-
> timately derived from active experience of those
> aspects of reality to which the words belong. The
> chemist or the physicist understands the meaning of
> his most abstract concepts ultimately on the basis
> of his acquaintance with chemical and physical processes
> in the laboratory. Even the pure mathematician, dealing
> with that most useless and arrogant branch of his learning,
> the theory of numbers, has probably had some experience of
> counting his pennies and shillings or his boots and buns.
> In short, there is no science whose conceptual, hence verbal
> outfit is not ultimately derived from the practical handling
> of matter. I am laying considerable stress on this
> because, in one of my previous writings, I opposed
> civilised and scientific to primitive speech, and
> argued as if the theoretical uses of words in modern
> philosophic and scientific writing were completely
> detached from their pragmatic sources. This was
> an error, and a serious error at that. Between the
> savage use of words and the most abstract and theoretical
> one there is only a difference of degree. Ultimately
> all the meaning of all words is derived from bodily
> experience. (quoted by Langendoen 1968, 34)

I can now turn to what I feel to be the philosophical
speculation most relevant to the linguist's business, namely
"the alleged universality of semantic components" (Lyons
1968, 472-5) just referred to in connection with the claims
made by Katz and his associates about their "semantic markers",
and other similar claims. As Lyons notes, it "has frequently
been suggested that the vocabularies of all human languages
can be analysed, either totally or partially, in terms of
a finite set of semantic components which are themselves
independent of the particular semantic structure of any given
language" (ibid.). Cherry, for instance, discusses at some

length the proposals made in this connection by such authors
as Leibnitz, Descartes, Dalgarno, Bacon, Spencer, Ampère, and
others (Cherry 1957, 37-38 and 228). While Katz and his
associates have concentrated on such problems, Chomsky's
contribution in this respect is limited to a few general
assumptions which he leaves unexplored : for instance,
"that the color words of any language must subdivide the
color spectrum into continuous segments; or the condition
that artifacts are defined in terms of certain human goals,
needs, and functions instead of solely in terms of physical
qualities" (Chomsky 1965, 29). In this connection, it may
be interesting to refer briefly to the work of Leech, not
so much because his is the only notable study of English
(next to those of Katz and associates) in terms of semantic
components, but especially because his work embodies a
metatheoretical shift directly relevant to the present
discussion. In a study of advertising English published in
1966, Leech states that to "describe the English of advertis-
ing is to say what is distinctive about it : that is, how
it differs from other 'Englishes'" and that only a "brief
consideration of regional and national differences is enough
to dispose of the myth of the English language as a total,
invariant system of rules" (Leech 1966, 67-8). In 1969,
however, Leech outlines a "semantics of English" in which,
he argues, "each particular descriptive problem" must be
(i) "related to a general semantic theory", (ii) "to other
'semantic fields' within the same language, so that features
of meaning may be generalised to the language as a whole",
and (iii) "to the relevant lexico-grammatical structures, so
as to form part of an integrated total description of the
language" (Leech 1969, v). I suggest that the dilemma
between "distinctive" and "total" description of the language
can be clarified by distinguishing those aspects of it that
are, or can be, a matter open to research in semantic or
linguistic theory from those that fall within the realm

of philosophical speculations - were it only because they
are in contradiction with available facts. Under this view,
I suggest that there is no conflict between the distinctiveness
requirements as discussed in the preceding section and such
attempts as have been made by Leech 1969 and others to
"generalize" questions like 'Time', 'Place', and 'Modality'
in English "to the language as a whole". What I wish to point
out, however, is that the universality of semantic markers
or components, and their validity for the whole of a language,
as well as the fiction of a universal semantic language,
however interesting it may be, are still a matter of philoso-
phical speculation : "such empirical evidence as there is
available at the present time would tend to refute, rather
than confirm, this hypothesis" (Lyons 1968, 473). The
evidence is particularly clear in the field of documentation.
In this field, the structural unity of the vocabulary of
"universal" classifications has to my knowledge never been
shown, on the contrary (Coyaud 1966). As I shall show below
in connection with the field of the abstracts under investiga-
tion in this thesis, the problem is rather the diversity
of existing documentary languages or classifications, even
in a single field. A similar diversity appears in the existence
of various documentary languages serving as metalanguages
for the various scientific disciplines and in the practical
impossibility of reducing them to a cross-disciplinary scheme,
or again, in the commonly accepted differences between
competing linguistic theories, as usual practices in current
linguistic research show (Botha). In the same fashion, Gardin
notes the following, in his introduction to Allard et al.'s
conceptual analysis of the Coran :

> the orientation of research and, consequently, the way
> in which observations are differentiated are indeed
> quite different according to whether one deals with
> chemical reactions, recipes, or philosophy. The mental
> processes involved are no doubt akin; but the organiza-
> tion of the experimental field, in some way or other, by
> means of the vocabulary, varies to such an extent that
> it seems unreasonable to try to reduce it in each case
> to a single mould (Allard et al. 1963, II 18; my trans-
> lation from French).

Similar points have been made in the linguistic literature
on several occasions. Bolinger has suggested that katz and
Fodor's lexical entries are unable to incorporate various
types of information about "universes of discourse", and
katz and associates have not to my knowledge made any proposals
to solve the problem. Bolinger, for instance, is doubtful
about how katz and Fodor's semantic representation of lexi-
cal items could serve to disambiguate the sentence He became
a bachelor in 1965, by incorporating the information that
"knighthood died out a long time ago" (Bolinger 1965, 561).
As he also points out .

> The radical shifts effected by metaphor throw into relief
> the bold differences between universes of discourse in
> the totality of language. They raise the question of
> whether it will ever be possible to attach everything,
> from slang to technology to science and back again
> to politics, religion, and the home, with one arsenal
> of categories. (ibid., 567)

Harris has recently made a similar point, from the
angle of the distributional linguist, for whom the basic
fact to account for is the extent to which some sentences
are "acceptable " :

> many sentences are really to be found only in particular
> types of discourse, i.e., in the neighbourhood of parti-
> cular other sentences ; and many of these would indeed
> be but dubiously acceptable sentences outside of such
> discourses. Thus The values approach infinity is normal
> in mathematical discourse but rather nonsensical outside
> it. It rotates about the bond, acceptable in chemistry,
> may be nonsense in ordinary English (...) If meteorites
> flew down all round us is an acceptable sentence, it is
> not because this is a meaningful word combination, since
> flew being here not "flew" in the usual sense but some
> movement distantly similar to it. The datum is simply
> that flew is used here, i. e. occurs in an acceptable
> sentence with meteorite, by virtue of some extension
> or other process of language use. (Harris 1968, 52).

The solution, then, which does not commit the linguist to
philosophical speculations, consists of treating the diversi-
ty of metal nguages as a fact by giving "theoretical recogni-
tion" in linguistics "to the tacit presuppositions and

assumptions in the speech-community" and to take "no account
of their validi   within some   other frame of reference
assumed to be absolute or linguistically and culturally
neutral" (in a discussion referred to above on synthetic
vs. analytic truth and on "pragmatic implication" : Lyons
1968, 445; cp. Fillmore 1967 : "The Grammar of HITTING and
BREAKING", studied in their "non-transferred meanings";
and Fillmore 1969b on 'Verbs of Judging').

In contrast with such a position, Katz and Fodor
believe that when people are unable to produce or understand
"sentences with technical words" which they don't know, such
"exceptions" to the fluent speaker's "ability to produce
and understand any sentence of his language" are "of no
systematic importance" (1964, 481). In the present state
of linguistic semantics, I believe this to be a gratuitous
claim, involving a philosophical commitment to be rejected
as in contradiction with facts (cp. Harris's related, one-
page programmatic proposal to "conjoin" sublanguage meta-
languages with each other by means of metalinguistic
conjoined sentences, in Harris 1968, 155). The purpose of
what follows, then, is not to present any kind of psycholog-
ical hypothesis or philosophical speculations, but to
introduce facts that will characterize the present study -
despite the occasional use of such convenient terms as
"concepts", "knowledge of the world", etc., according to a
common practice in linguistics and other branches of language
studies.

1.1.6. Semantic analysis :
its components and formal requirements.

I now wish to outline the metatheoretical conception of
semantic analysis which underlies the present thesis, after
a recent comparative study by Gardin of "semantic analysis
procedures in the sciences of man" (Gardin 1969). In order
to do this, it is necessary to refer back to the discussion
on types of metalanguages presented in section 1.1.4. In
this discussion, I suggested that a consideration of the
communicative function of language imposes a distinctiveness
requirement on semantic analysis and that such a requirement
can be met by selecting the appropriate type or types of
metalanguages. I also suggested that to be rightly called
a linguistic theory, a metalanguage needs to be describable
as a 'public' metalanguage, and I have argued that the
documentary languages such as classification systems, which
are used by documentalists, meet this description of 'public'
metalanguage : both linguistic theories in this sense and
classification systems or the like are set up and used by
a community and can thus perform a communicative function
which is not performed in the same way by what has been
called 'private' metalanguages, and illustrated by the
semantic markers used by Katz and the definitions underlying
them. As I have also pointed out, the difference between
a linguistic theory and, say, a classification is that a
classification is 'field-specific'. A classification is set
up for a group of people or observers to make distinctive
semantic statements in a field, while a linguistic theory
does not serve this function : no linguistic theory, for
example, can serve to make semantic statements that are
distinctive and 'public' about English abstracts in the field
of documentation, or, more precisely, about the English of
documentation abstracts - the object-language under scrutiny
in this thesis.

The above distinctions are, I believe, intuitively
clear and rest on unquestionable facts - at least as far as
the topic of this thesis is concerned : even if the situation
is different (and, as I have argued, obscured) in the case
of semantic analysis of 'everyday' language, this does not
affect the point which I am making, and which could equally
be made about a host of other 'fields'. The purpose of
this discussion, however, is not to show how clear my dis-
tinctions are. On the contrary, these distinctions are
intended to introduce the fundamental PROBLEM of this thesis,
as I see it after suggestions made in Gardin's discussion
of semantic analysis as practised in particular by the content
analysts of the Stone group, and by documentalists such as
those working around Gardin himself and a few others.

After Gardin 1969, the content analysts, and recent
research in TG theory, I shall now make use of the following
terminology :

> (i) Metalanguage will stand for what can be described
> as 'field-specific', 'public' semantic metalanguages :
> documentary languages such as classifications, or the
> content analysts' systems of "characteristics" or
> "categories" such as those illustrated above with
> 'conservatism'. The metalanguage used in the present
> study is a kind of classification set up for the field
> of scientific information (documentation) by N. Gardin
> and F. Lévy.(1964).
> (ii) Theory or linguistic theory (rather than the term
> model used by Gardin 1969 and applied to non-linguistic
> models) will serve to refer generally to theoretical
> research in linguistics or, more specifically, to the
> selection made in this thesis of those theoretical
> points in linguistics that I believe to be directly
> relevant to my topic.
> (iii) Procedure is the term which, after Gardin, I
> shall apply to attempts made in the fields of content
> analysis and documentary analysis to express the rela-
> tions established empirically between a given meta-
> language and a given object-language in the form of
> a system of rules meeting certain requirements of form
> discussed below, and amenable to mechanization or
> computerization. As already noted, Part II of this
> thesis presents such a procedure, which I have set up

and computerized at the University of Victoria.
This procedure relates the classification system mentioned
above in (i) to a few English abstracts in the field
of documentation. As far as linguistic theory is concerned,
my procedure is an attempt to make specific statements
after general rules such as those proposed in Harris's
string analysis theory. (Harris 1965).
(iv) Deep and surface structure. Unless otherwise
specified, I shall use these terms not in the meaning
in which Chomsky 1965 and others after him have used
them in "syntactic' theory but rather in the sense in
which some of Chomsky's disciples such as Lyons,
Fillmore and Langendoen, use these terms to make a
broad distinction between the semantic representations
of sentences (their deep structures)and the representa-
tions of them which are closest to their physical
manifestations (hence 'surface' structures), for instance
as words on the page as is the case of the written
material under investigation in this study. The
distinction as used here, however, can be said to
originate from a problem raised by Chomsky himself.
In a passage which may serve as an illustration of the
kind of distinction I shall be concerned with, Chomsky
1965 (162) notes that "there are cases that suggest the
need for an even more abstract notion of grammatical
function and grammatical relation than any that has been
developed so far, in any systematic way". Chomsky notes,
for instance, the "meaning relation, approaching
a variety of paraphrase" that holds between sentences
such as "I liked the play","the play pleased me".
To these, one could add other equivalent sentences
such as I liked what they performed, and what they
performed pleased me. A deep structure representation
of such sentences, then, will have to state somehow
the equivalence in meaning between them, while their
surface structure representation will account for the
way they actually appear as written or spoken material.
For instance, the surface structure representations
of the above sentences will state that like requires I,
and please requires the play as subjects, and will state
the functional similarity between the play and what they
performed as subjects of please. Ideally, a linguistic
theory such as that currently under investigation by
Lyons, Fillmore, Langendoen and others will also relate
both types of representation to each other by rule. This,
however, cannot be done in the present state of linguistics
on the basis of a generally accepted theoretical frame-
work : typically, Leech relates "semantic and grammatical
structures only informally" (Leech 1969, vi), and Fillmore
has refrained from publishing transformational extensions
of his work on semantics ever since 1968. The only
thing, then, that seems to be generally accepted within
the distinction between deep (semantic) and surface

60.

(grammatical) structure is that semantic and grammatical
descriptions should use different symbolic conventions.
As I shall point out later, those proposed for semantics
are based on the quasi-logical predicate-argument form,
while those used for surface structure descriptions are
based on more traditional-looking symbols of grammar
such as those made familiar by Harris or by Chomsky
(Noun Phrase, Verb Phrase, Noun, Verb, etc.), on the
basis of which grammatical functions or relations such
as subject-of can (but need not) be defined.

I am now in a position to outline the metatheoretical
proposals made in this thesis. After Gardin 1969, I propose
to regard metalanguage, procedure, and theory as "the three
components of semantic analysis" (Gardin 1969, 34) and to
consider the purpose of each particular semantic investiga-
tion as that of "clarifying" the nature and interrelations
of these components, by meeting certain requirements of form.

The following chart may serve to present my
metatheoretical proposals and the general organization
of this thesis :

METALANGUAGE

(a classification
for the field of
documentation) :

PROCEDURE :

OBJECT-LANGUAGE

(the English of
abstracts on
documentation) :

deep structure	surface structure

deep structure	surface structure
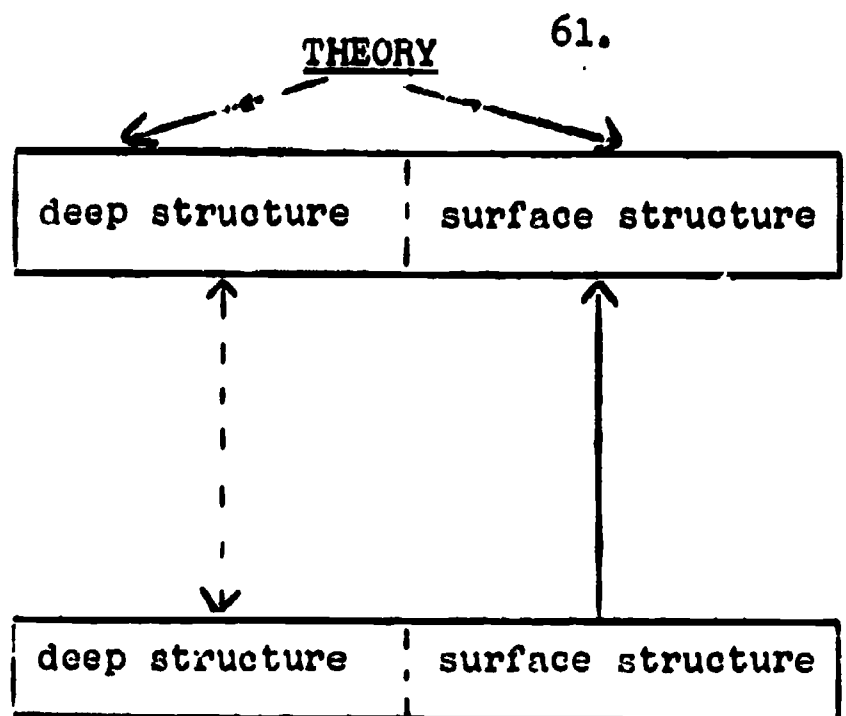
<u>**Figure 1**</u> · <u>**the components of semantic analysis.**</u>

The dotted lines and arrows indicate what I believe should
ideally be done but can only be tentatively investigated in the
present state of language studies, while the continuous lines
represent what has actually been accomplished, and computerized
in Part II of this thesis. In the computerized procedure which
I present, each of a set of fifty English abstracts in the
field of documentation has been reduced mechanically to one
or more entries of a classification for this field. As
symbolized in Figure 1, the rules used by the computer treat
the abstracts analyzed as a succession of sentences, and
these rules specify the <u>surface</u> structures of these sentences.
The classification entries into which the abstracts are
eventually reduced are equally treated as surface structures
of the general form "<u>X which is defined as...</u>", corresponding to
the fact that the classification system used consists of 'words'
with definitions.   ~~with~~

As noted earlier, Harris's _string analysis_ theory provides
the main basis for the surface structure analyses presented
in this thesis.

Part III of this thesis is devoted to deep structure
considerations, i.e. to problems of semantic representation.
I argue that these are relevant to both the object-language
(that of the abstracts under investigation) and to the
classification entries used as a metalanguage. About these
entries in particular, I shall assume that, since they have
the form of 'words' with definitions in a natural language,
the theoretical distinction between deep and surface structure
needs to apply to them, as it does to the object-language
data. On the other hand, their metalinguistic status also
needs to be given theoretical recognition. To take an extreme
example, such a recognition will enable us to ignore the
relation, which a "detached linguist-observer" might like to
investigate, between the meaning of _retriever_ as 'dog' and
the meaning of the same symbol in an abstract of the Janaske
corpus : "the individual as a user, generator, and _retriever_
of information" (Janaske 82). The two-directional arrow
written in dotted lines in the above chart symbolizes my
assumption that an ideal procedure for documentation purposes
would relate two kinds of deep structures (one for a meta-
language such as a classification and the other for, say,
abstracts in a particular field) and that it would do this
both ways, i. e. that it would be _reversible_. This means that
ideally a computer system for semantic analysis could both
analyse the texts of the object-language into their deep
structure representations and these in turn into deep structure
representations of the metal nguage, and similarly synthesize
surface structures of abstracts using surface and deep
structures of the metal language as input. As Gardin 1965
has argued, a reversible procedure could be used as a means
of experimental verification, by a calling to the judgments
of competent informants (Gardin 1965, chapter 5).

The role proposed for the procedure in this study is
ideally to make the empirically established correspondences
between object-language and metalanguage "explicit" or
"formal" by accounting for them in terms of operations
stated in the form of rules. As already noted, the part
played by the theory in setting up these rules is a matter
of investigation. As Gardin puts it, an ideal procedure

"must be precise and complete : precise, in the sense
that the analytic operations must be explicitly defined,
and complete to the extent that these analytic opera-
tions must adequately account" for the passa e from
object-language to metalanguage, and vice-versa.
(Gardin 1969, 18).

The requirement of explicitness has already been defined
above after Stone et al. (1966, 11) : an explicit procedure
is one that can be "replicated exactly by other analysts".
The symbolic aspect of this explicitness requirement is an
aspect which is directly relevant to the role of the computer.
It can be described in terms of Katz and Fodor's ideal of a
"formal" statement as one "in which the application of rules
is defined solely in terms of the shapes of the symbols
they apply to, and the operations which the rules effect
in producing their output are mechanical", i e. do not rely
on "the speaker's intuitions or insights" (Katz and Fodor
1963, 192).

While I shall have to return to Chomsky's distinction
between competence (broadly, theoretically defined "knowledge
of the language") and performance (roughly, actual use made
of this knowledge), I must note at this point that the procedural
conception of semantic analysis outlined here is presumably
in opposition with views expressed by Chomsky and some of
his associates (e. g. Katz and Postal 195 , 166).
I shall assume that Chomsky uses the term procedure in uch
the same sense as I have done here, in spite of a passage
where he defines a "reasonable procedure" as one that does
not involve extralinguistic information -- that is, one that

does not incorporate an "encyclopedia."* (Chomsky 1965, 201-2, ftn. 17). For Chomsky, then :

> the grammar does not, in itself, provide any sensible procedure for finding the deep structure of a given sentence, or for producing a given sentence, just as it provides no sensible procedure for finding a paraphrase to a given sentence. It merely defines these tasks in a precise way. A performance model must certainly incorporate a grammar ; it is not to be confused with a grammar. (ibid., 141).

If Chomsky's distinction between a "procedure" and a "precise" definition of the tasks involved in one is intended to emphasize the 'state of our ignorance' I can only agree. If, however, it is to be read, after Leech, as the expression of a "rather novel philosophy ` science" according to which "the primary purpose of science is not to explain and predict observable happenings, but to "gain insight" (Chomsky 1965, 20)" (Leech 1968, 90) I can only disagree.

1.2. METALANGUAGE

1.2.1. Compatibility between metalanguage and object-language.

This section deals with my selection of the object-language and of the metalanguage for this study. As will be remembered, I wish to take for granted the general suggestion made by Leech to regard the semanticist's task as that of explaining what underlies statements such as x means y. I shall touch upon some aspects o Leech's suggestion below, under the headings 'Theory' (1.3.) and 'Procedure' (1.4.). Here, I wish to return to Leech's suggestion in one of its aspects which is relevant to a central fact about the present investigation. After Leech the purpose of this study can be described as that of accounting for what "underlies" (in a sense to be explained later) semantic judgments which are statable as an extremely complex sentence of the type "I (the linguist observer actin as an expert-indexer) relate x to y." In this sentence, x is an abstract belonging to the object-language, and y consists o one or more entries of the classification chosen as a metalanguage. The object-language is the English of abstracts on documentation, and I drew these abstracts from a specialized bibliography compiled by an American team and edited by P. C. Janaske (Janaske 1962). The classification used as a metalanguage has been set up by J. Gardin and F. Lévy (1967). As will appear are the following example, the sentences by means of which my semantic judgments can be expressed are extremely complex, as x is a whole text (an abstract) which is posited to be reducible to a sentence (see 1.3.2.) and y, one classification entry (or more), consists of a noun phrase with a definition : the original version of the classification is in French and my quotations from it will be given in my own English translation. Consider, for instance, my semantic judgment about abstract 1 of the Janaske bibliography :

<u>I relate</u>

(the abstract)

The Japan Information Center of
Science and Technology and its
activities.

A general description of the Japan Information
Center of Science and Technology which was es-
tablished for the purpose of contributing to the
development of science and technology by collect-
ing, processing, storing, and retrieving scientific
information. At the time of this report, the Center
was planning to change over to a mechanized storage
and retrieval system to that used by the U.S.
Patent Office. Services available from the Center
and the charges for each are listed.

similar ✓

<u>to</u>

(the classification entry)

<u>Scientific and technical information institution</u>

<u>which is defined as :</u>

<u>institutions more or less directly concerned with scientific</u>

<u>and technical information (archives, indexing services,</u>

<u>translation centres, etc) when studied individually -</u>

<u>structure, functioning, cost, budget, etc. - or by types :</u>

<u>role, models, evolution, in particular with respect to</u>

<u>mechanization.</u>

Figure 2 : Example of a semantic statement.

The problem I wish to raise here is on what basis it can be assumed that the abstracts of the Janaske bibliography and the Gardin classification are compatible, or relevant to each other. In other words, I wish to raise with respect to this study the question referred to above as that of observational appropriateness conditions, and illustrated with the case of book titles ("Advanced microbiology" versus "x Advanced microbiology for the elementary school"). This question is best answered by again resorting to Leech's suggestion. The title of the Gardin classification (which includes 25 main entries such as that quoted above) explicitly mentions that the classification is "in the field of scientific and technical information". Besides, the Janaske bibliography is entitled "Information handling and science information. A selected bibliography 1957-1961". My assumption, then, concerning the compatibility between the two can be stated in the form of the following judgment, which relies on the experts' judgments just quoted and posits a semantic equation between the two :

I assume that

the Gardin classification

which is said to be "in the field of scientific and
technical information"

(more conveniently referred to as "documentation")

and

which consists of the following 25 main entries : "..."

is relatable to

the Janaske bibliography

which is entitled "Information handling and science information"

as explained in the following preface (...)
and
which consists of the following abstracts : (in total 1121)

Figure 3 : Statement of the assumed
compatibility between the
object-language and the
metalanguage.

This is an obviously huge and complex sentenc., which
needs to b. somehow reduced to more researchable proportions.
I have given it here, however, for several reasons. First
of all, it embodies a fact the complexity of which comes close
to what TG linguists have been able to imagine on the basis
of their highly abstract conception of the sentence :
sentences too long to be processed in one's lifetime, etc .
Secondly, the situation which it symbolizes is even more
complex than might at first appear if one considers it in
the light of Chomsky's notion of competence, or rather a
version of it which I attempt to define below (1.4. 1.). In
brief, the problem is not so much that people actually
"perform" the action of uttering an "appropriate" judgment
involving 25 classification entries and 1121 abstracts. The
problem lies rather in defining their 'competence" or ability
to utter such statements or judgments, including some that
can involve far more than the 11.1 abstracts compiled by
Janoska. Concretely speaking, it is the problem of why an
expert in documentation working in a library or documentation
centre can use the Gardin classification (or another : see
1.2.2., below) to represent the meanings of any abstract or
text in his field, while he has never come across them before.
Thirdly, and perhaps most importantly as far as my proposed
theoretical conceptions are concerned, Figure 3 and the
passage from the Janoska preface quoted presently point to
a puzzling paradox related to that of field-specific meta-
languages mentioned above : it seems to be easier for people
to utter semantic judgments involving a huge and complex
mass of data than to utter specific judgments concerning parts
of these data. This appears most clearly from the fact that
the compilers of the Janoska bibliography were unable to
index each individual abstract of their bibliography (they
merely list the abstracts alphabetically by author) while they
seem to be perfectly confident in their pertinence judgment
concerning the bibliography as a whole.

The following passage from the preface to the Janaske
bibliography provides evidence for what procedes. Besides,
it points to various intricacies of the field of "scientific
and technical information" upon which I shall not comment,
and to the crucial fact that the Janaske corpus represents
only an inconsiderable portion of the indeterminate number
of abstracts and texts "on the subject of information handling
and science information" :

Conducting investigations ar . doing research problems in
the area of science information requires literature support
just as any other research. The sources of information
and the organization are somewhat more complex and
difficult to manage. The "field" is new, and a communica-
tion problem is present in every field of interest.
There is no single source one can consult to locate and
identify reports ranging in scope from data handling to
information storage and retrieval; or from classification
schemes to the administration of a document collection.
Reports of work falling within the scope of interest may
be found in trade and industrial reports, government
reports, and a wide variety of serials ranging from trade
and commercial to specialized journals. It is interest-
ing to note that nearly every indexing service includes
the terms "automation," "communication," and "documenta-
tion," but only a small percentage of the entries are
pertinent to the   t of information handling. Collect-
ing reprints or   bibliographical references relating
to information in this field of interest very soon
produces a large volume of paper. It is unrealistic to
confine one's efforts related to information handling
to a single subject discipline - the basic principles
or routines which may be applied to the chemists may be
equally useful and adopted for use by the biologists.
The subject of "information handling and science infor -
mation" as used here covers the full range of science.
The arrangement of the bibliographic entries is alphabeti-
cal by author. (...) Each item has been numbered for
indexing purposes. The major emphasis in making the
selection of items to be included in this bibliography
is handling of scientific information and the application
of mechanical systems of information storage and retriev-
al. In addition to this, related areas of interest such
as scientists' use of literature, guides to the use of
scientific literature, mechanical translation, training
of information specialists, reprography, subject classi-
fication, and indexing have been included. In order to
limit the size of this bibliography to a manageable pro-
portion, the past five years (January, 1957, to December,
1961) is the period represented by the items in this
bibliography. A few items which have appeared early in
1962 have been included because of their significance.

One last and even more difficult point needs to be mentioned here. One may wonder why I have selected a bibliography of <u>abstracts</u>, rather than integral texts, for the present investigation. The tentative answer is this. Beyond the intuitively obvious point that abstracts will be less complex than the corresponding integral texts, it must be assumed that the available classifications in the field of "documentation" are - in some obscure way - <u>less</u> compatible with integral texts than with abstracts, in that they do not define a "culture" rich enough to deal with integral texts. Intuitively, this even seems to be the case for some abstracts of the Janaske bibliography, as the following example suggests :

> 551 KREMENEKAJA, A.V., and LEV, D.I.
> New methods for presenting biblio-
> graphical information in the U.S.S.R.

Two methods of disseminating scientific and technical articles are described as used by the State Public Scientific and Technical Library attached to the Siberian Department of the USSR Academy of Sciences. The first method, which is suitable only for individually prepared bibliographies or other cases requiring only a few copies, involves the use of pockets the size and shape of catalog cards into which microfilm copies of articles are inserted. The bibliographic description, classification number and card number are printed on one side of the pocket, and the serial and classification numbers are shown on the first frame of microfilm. Production costs are about two rubles a card. For current information on a large scale, microcards are used. Since the average article is only four pages in length, a modified type of card is prepared which reduces the original only five or six times. The cards can, therefore, be read with very simple magnifying glasses. Bibliographic information is photomechanically printed on a catalog card, on the back of which is pasted the photographic reproduction of the original. At the beginning of 1960 the Library opened a subscription list and inaugurated a pilot service of this kind for mechanical engineering and plastics.

1.2.2. The notion of "concordance".

As Botha has noted, it is a fact that there exist in linguistics, like in other disciplines, several competing theories, and the relations between them or between their constituent concepts are far from clear or straightforward. In this section, I wish to point out that the same is true of metalanguages (documentary languages) for a single field, that this is a fact which needs to be given recognition, and that documentalists (and to my knowledge, nobody else) have solutions to propose with respect to this fact. First of all, I wish to concur with Gardin's view that "the diversity of documentary languages" is :

> the reflection of the necessary multiplicity of points of view on the empirical world, not so much for the ideological reasons which instinctively come to mind (freedom of thought, etc.) as for scientific reasons which are much more serious. These stem from the fact that one has to develop different symbolic quantifications in regard to experimental phenomena of the same order, so as to build up complementary theories about them (...) That all these theories may ultimately be reconciled - there is nothing to prevent one from thinking or believing so and hence to speculate on an ideal symbolization compatible with each specific scientific language (pertaining to a special field) - but one must always bear in mind the paradoxical characteristic of the temporal success of any such project : it would imply that scientific research has come to an end, having reached its ultimate goal. (Gardin 1969, 22-23).

Such a position has, I suggest, received metatheoretical recognition in various forms from some students of language, but has been given astonishingly little theoretical attention. Thus, in a discussion of implication, Lyons stresses that understanding is relative to the observer or addressee :

> What is normally meant by 'understanding' utterances can be quite well accounted for without making the assumption that all speakers of a language will draw from an utterance exactly the same set of implications. (Lyons 19 , 44).

Lyons's position can be compared with Harris's proposal
to analyse sentences in terms of reconstructed 'metalinguistic"
conjoined sentences, and with his insistence that these recon-
structions "will vary for different audiences" (e.g. They
described Vesuvius A D VESUVIUS IS A VOLCANO rather than a
mountain : Harris 197 :16C). This, it should be noted, is
different from the conception of Katz and Fodor that "any
sentence may be made to mean nything you like simply by
constructing the setting to include the appropriate stimula-
tion" (Katz and Fodor 1964, 4C8) . here, the only communication
situation envisaged is (at best) that in which the linguist-
observer acts as sole addressee, and only his personal intuition
is confronted with the full complexity of the conceptual inde-
terminacy of linguistic utterances in the whole of a language.

In order to face the practical problems resulting from
the existence of several documentary languages in a single
field or discipline, documentalists or information scientists
have developed "compatibility" or "convertibility" studies
(Henderson et al., 1966, in Gardin 1969). I shall not
attempt to discuss these studies here. What matters for the
present purpose is that they can lead to the setting up of
what may be called, after J. Gardin and F. Lévy, "concordances"
between several classification systems in the same field ;
other names for such conversion systems between classifica-
tions and the like are "intermediary lexicon" or "basic
lexicon" (Gardin 1969, 23). Gardin and Lévy's metalanguage
for the field of documentation is such a system. A description
of it can be found in N. Gardin 1969. In brief, it has been
set up to ensure convertibility between essentially 6
classifications from several countries and institutions.
Moreover, I must stress the fact that the concordance results
from a close cooperation (meetings, discussions, etc.) between
N. Gardin and J. Lévy on the one hand, and the authors of
the classification systems involved, on the other. This fact

points to the _public_ char cter of the Gardin concordance,
in the sense of my discussion on types of metalanguages
(1.1.4.).

In spite of the fact that such concordances"are not
intended to replace existing information retrieval languages"
(i.e. documentary languages : Gardin _ibid._), I have decided
to make use of the Gardin-Lévy concordance as a metalanguge
for this study because it enabled me to achieve an important
preliminary simplification of my research problem. As the
concordance is organized around 25 main headings with
definitions, I decided to use these 25 headings as a meta-
language, thus reducing the problems raised by my mechanized
indexing experiment (Part II) to a study of the relations
between a limited number of abstracts and the 25 headings
in question. The headings (called "groupes") together with
their French definition are quoted in Appendix 1. My
English translation of the headings is given in Appendix 2.

1.3. THEORY

1.3.1. 'Sentence', deep and surface structure : an evolution.

This introductory section is intended to outline a
fairly recent evolution (if not a revolution) which has
taken place in TG theory, the linguistic theory initiated
by Chomsky mainly after Harris's proposals on "transforma-
tions". Here I am less concerned with this evolution as such
than with some of its developments because th se form the
basis of the theoretical framework I wish to propose in this
study. My discussion will proceed in four main stages. First,
I shall briefly illustrate the theory proposed by Chomsky
and his associates, around their distinction between deep
and surface structure and around the conception of the
semantic component of this theory as proposed by Katz and his
associates, and the parallel metatheoretical indeterminacy
(the indeterminate boundaries separating the domains of
syntax, semantics, and "systems of knowledge and belief" in
Chomsky 1965). Secondly, I shall consider recent alternative
proposals, which are at least partly intended to put an end
to this indeterminacy. Thirdly, I wish to point out why, in
my opinion, Fillmore and a few others go a step further in
the same direction. Fourthly, I shall examine the evolution
of the notions of "sentence" and "sentence context" within
the general metatheoretical and theoretical evolution under
discussion. From the outset, I must point out that this
discussion reflects my own interpretation of the developments
of TG theory in the last five years or so, as a history of
these developments has not to my knowledge been written so
far (on this topic see my recent paper : J. Noël 1971).

Using the terms "deep" and "surface structure" apparently
first introduced into linguistics by Hockett, Chomsky has
based his theory of syntax on the "central idea" that

> the syntatic component of a grammar must specify, for
> each sentence, a <u>deep structure</u> that determines its
> semantic interpretation and a <u>surface structure</u> that
> determines its phonetic representation,

that the deep structure and the surface structure of a
sentence "are, in general, distinct" and that "the surface
structure is determined by repeated application of certain
formal operations called "grammatical transformations""
(Chomsky 1965, 16) to deep structure representations.
As Botha most succinctly illustrates it :

> To take an example from Chomsky (1964a, p. 34), in the
> representations of the surface structures of the
> sentences
> (7) (i) <u>John is easy to please</u>
>   (ii) <u>John is eager to please</u>
> it is not specified that while John is in 7 (i) the
> direct object of <u>please</u>, it is in 7 (ii) the logical
> subject of <u>please</u>. Secondly, in the case of syntactical-
> ly ambiguous sentences, more than one deep structure
> must be taken as underlying one surface structure. For
> example, to account for the ambiguity of
> (8) <u>what disturbed John was being disregarded by</u>
> <u>everyone</u>
> two deep structures must be taken as underlying this
> one surface structure. (Botha 1968, 25-26)

Surface structure, then, is the grammatical representation of
sentences that is closest to their spoken or written form,
while the deep structure representation(s) is (are)
'abstract', i.e. removed from the surface structure, to a
degree to be determined by the theory. The two main
assumptions of the theory are that deep and surface structure
"are, in general, distinct" and that deep structure repre-
sentation is primary in accounting for people's ability to
express "indefinitely many thoughts" and to react "appro-
priately in an indefinite range of new situations" by means
of the sentences they produce and understand (Chomsky 1965,
5 : this, as will be noted again later, is a fundamental
point in the TG conception of language).

As already noted, Katz and his associates developed a
semantic theory to serve as the semantic component of
Chomsky's general theory, and this component was to
operate on the deep structure re₁ sentations of sentences.
For this reason, it has been called "interpretative", in the
sense that it interprets semantically the (deep structure)
output of the syntactic component of the theory.

For Chomsky 1965, however, as well as for Katz and
his associates, the boundaries separating syntax and
semantics indeterminate, that is, the question of
determining the respective domains of the syntactic and
semantic components of the theory is a matter of investigation.
This appears most conspicuously in Katz and Fodor's famous
negative delimitation of semantics : "synchronic linguistic
description minus grammar equals semantics" (1963, 172).

In the mid-sixties, an evolution in TG linguistics took
place which can be described in terms of the assumption
"that there is no real distinction between syntax and se-
mantics" and "that the rules for combining material into
sentences are really rules for combining semantic material".
(McCawley 1968, 586). The obvious consequence of such a
conception is that the existence of syntax and semantics
as distinct components of the theory may be unnecessary.
This is particularly clear in Fillmore's comments on his
motivation for setting up his theory of semantic deep
structure relations ("case" relations) :

> If it is possible to discover a semantically justified
> universal syntactic theory along the lines I have been
> suggesting, if it is possible by rules (...) to make
> these 'semantic deep structures' into surface forms of
> sentences, then it is likely that the syntactic deep
> structure of the type that has been made familiar from
> the work of Chomsky and his students [will appear as...]
> an artificial intermediate level between the empirically
> discoverable 'semantic deep structure' and the observa-
> tionally accessible surface structure, a level the proper-
> ties of which have more to do with the methodological
> commitments of grammarians than with the nature of human
> languages (Fillmore 1968a, 88).

While the conception as formulated by Mc Cawley makes
it possible to quibble by saying that, after all, one's
concern with "syntax" has always been a concern with seman-
tics, the reality of the evolution appears unequivocally in
the above quotation from Fillmore and in other recent studies
- in the topics they deal with and in the kind of problems
they raise and arguments they use. A paper by Lakoff (1968b)
is another good example of a radical departure from the TG
as well as other traditions. In this paper, Lakoff advances
a number of arguments for considering that sentences of the
types <u>x did something with y</u> and <u>x used y to do something</u>
have the same or a similar deep structure. While he does not
say what this deep structure would be like, the suggestion
is that it is needed in the theory in order to avoid having
to state again and again the same rule for a number of con -
st ructions in which the two types of sentences clearly have
the same 'semantic' and 'syntactic' properties (e.g. "<u>x</u> must
be animate").

I now wish to argue that such researches as Fillmore have
carried the evolution under discussion even a step further.
In a paper written as early as 1965, Fillmore was concerned
with what he then called "entailment rules". Typically, he
explicitly presents these rules as rules of semantic inter-
pretation which lie <u>outside</u> the scope of the semantic inter-
pretative component of the theory set up by Katz and associates:

> entailment rules can be described, roughly, as operating
> in the following way : There is a sentence X which cannot
> by itself be interpreted by the ordinary semantic rules.
> Based on the grammatical structure of X, the entailment
> rules will convert X into a set of sentences Y such
> that each of <u>these</u> sentences <u>can</u> be interpreted by the
> ordinary semantic rules. The semantic interpretation
> of the set of sentences Y, then, is provided as the
> semantic interpretation of the sentence X. (...) Consider
> the difference between the sentences "She reads Sanskrit"
> and "She even reads Sanskrit". The second tells us the
> same as the first, but it adds that this fact is somehow
> surprising. The contribution of the word "even" cannot

be explained, it seems to me, by assigning it semantic
features of the usual kind. I would say that the sen-
tence "She even reads Sanskrit" is to be 'factored',
so to speak , into two sentences, namely "She reads
Sanskrit" and "One would expect that she does not read
Sanskrit". (Fillmore 1965, 65-6 and 68).

The present outcome of the evolution in some current
TG research has been noted above, in connection with several
studies : Fillmore's "Grammar of HITTING and BREAKING", his
presuppositional studies such as that on "Verbs of Judging"
(Fillmore 1969), Green's analysis of sentences with such
words as "too" and "not ... either" (Green 1968), R. Lakoff's
study of conjunction (1970). In all these studies, grammatical
surface structure has become the interpretative level of
representation. More importantly, what this level interprets
is a type of semantic representation of which is required
that it should be specific or distinctive about 'concepts' :
as appears from Fillmore's recent research, the
conceptual interrelatedness of lexical items and their
conceptual contribution to sentence meaning have become
central preoccupations in the (primary) semantic level of
representation. Since my concerns here are metatheoretical,
I can ignore the problem of whether Harris's contributions
are "in" or "around" TG theory, and point out that Harris 1968
provides clear evidence that this author is going much the
same way as Fillmore and others, with respect to the require-
ment for linguistic description to be conceptually specific.
While I need not be concerned with the problem here, it is
interesting to note that, like Green, Harris even goes a
step further than Fillmore, by giving the same theoretical
treatment to idiosyncratic conceptual facts as to those
that can be regarded as part of "the language" or of a
"sublanguage". In the same way, Green discusses types of
"not ... either" sentences that may be said to reflect the
conception of the world of some mentally disturbed person.

Harris, then, argues that any metalinguistic "CS" (conjoined
sentence) having the "property of being known statements of
the dictionary or grammar, or of being common knowledge"
(i.e. specific knowledge shared by addressor and addressee)
can be zeroed and, conversely, reconstructed : thus, "The
war will start unless he enters the room" is only fully
acceptable if the communication partners are able to
reconstruct it as, say, "Since the invaders threaten war
unless the Prince of Cambodia comes to their conference room,
the war will start unless he enters the room" (Harris 1968,
132-4).

I would argue, then, that all this points to the emergence
of one or more linguistic theories aptly described as "syntac-
tico-lexicologico-semantics" (Elliott et al. 1969, vi),
in the sense that they are designed to be specific about
conceptual facts and that they do this by means of an
'integrated' conception of language structure. In this con-
ception, 'dictionary' and 'grammar' are no longer treated
in isolation, but both at the same time. Such a conception
has actually been stated recently by Fillmore and Lehiste,
in the form of some basic assumptions underlying their
research project. They describe their problem as :

> that of determining the nature of lexical information
> in a generative grammar of English. The ultimate,
> though probably unachievable, goal of such efforts
> is the design of a dictionary which will make avail-
> able to philosophers, foreigners, and computers, every-
> thing they need to know about the meanings, uses and
> forms of the lexical items in our language, and which
> will provide this information in a usable format.
> In a very real sense, all research in generative lin-
> guistics is relevant to questions of lexicology. Deci-
> sions on the form and explanatory scope of the theory
> of grammar determine in very direct ways the difference
> between the general or 'grammatical' facts about a
> language and the more specific or 'lexical' facts about
> its individual words. The rules for determining the well-
> formedness of the abstract underlying structure of sen-
> tences are rules which in the end specify acceptable
> combinations of grammatically organized lexical items.
> (Fillmore and Lehiste 1968, v).

A further aspect of the evolution discussed here will be examined below under the heading 'Structural semantics', in connection with an adaptation of some logical theories to semantic deep structure representation. All of these proposed adaptations cannot be compared and evaluated in the present study ; only some of them will be discussed in Part III of this thesis. What I wish to point out here is that these recent developments add a new assumption to Chomsky's conception of deep structure as, in general, distinct from surface structure and as the primary level of representation. The new postulate is that the symbolic conventions used in semantic deep structure representation are different from the grammatical symbols of surface structure representation. I shall accept this assumption in the present study. I must note, in this connection, that a similar assumption has been made by information scientists before it was made in linguistics, in particularly by Gardin and co-workers : as early as 1964 (Cros et al. ; cp. Allard et al., 1965), Gardin proposed and tested a form of semantic representation based on logical theories in a model (SYNTOL) for automatic storage and retrieval of scientific information. This concurrence between documentalists and linguists provides additional justification for the attempt made in this thesis to bring the two fields closer to each other.

My fourth point concerns a shift in the conceptions of the "sentence" and of "sentence context" within the evolution under discussion. The first thing to be noted is that much of the originality and interest (s far as this thesis is concerned) of TG theory, including Chomsky's, lies in the highly abstract conception of the sentence at the level of "deep structure". Chomsky has opened up the theoretical possibility of considering indefinitely long or complex

sentences ; for instance, sentences too long for them to be
spoken or heard in one's lifetime. Thus, in a _syntactic_ deep
structure ("base") rule such as that given in Langacker's
handbook (1967, 109), the symbol S (=sentence) is rewritten
as a conjunction of indefinitely many S's. To me, such a
rule suggests that, even in the theory of _syntactic_ deep
structure inherited from Chomsky, the existence of a clear
notion of "sentence context" is far from obvious. The same
uncertainty concerning this notion in a _semantic_ conception
of deep structure appears in Robin Lakoff's recent study
already referred to on the semantics of conjunction (R.
Lakoff 1970). In this study, it appears that Robin Lakoff
has ceased to distinguish sharply between a "sentence"
and a "discourse". I shall discuss this point in more
details in Part III, since R. Lakoff's paper is one of
the main sources upon which the semantic part of this
thesis is based. In connection with the difficulty of dis-
tinguishing between "sentence" and "sentence context", let
us also note the famous hypothesis of Katz and Fodor, which
is formulated in a passage quoted in the following section.

1.3.2. Surface structure of 'sentences' and beyond.

The conception of surface structure proposed in this
thesis is an adaptation of Harris's string analysis theory
(Harris 1965), more precisely of an extension of it by
Sager and co-workers (Salkof and Sager 1967). This extension
can be outlined around the notion of "restrictionless"
string grammar. The nature of my adaptations is determined
in part by computational requirements, but especially by the
role of a metalanguage in the metatheory proposed for this
study, as will be briefly shown presently. The chief claim
which Harris formalizes in his string theory can be described
as follows. Assuming that the contiguous "morphemes" or
words forming acceptable discourses can be represented as
acceptable sentences, these can be regarded as contiguous
strings consisting of contiguous "grammatical" classes
(parts-of-speech categories) and subclasses, and these
strings of contiguous classes and subclasses can be represented
by lists of axiomatic string formulas. These formulas, then,
state what sequences and dependencies of classes and sub-
classes are combined into acceptable sentences of a language
or sublanguage. An "acceptable" sentence is said to "satisfy"
the formulas in question. Harris's basic postulate for defining
string formulas can be described in Salkof and Sager's terms :

> Each sentence of the language consists of one elementary
> sentence (its center string) plus zero or more elementary
> adjunct strings which are adjoined either to the right
> or left or in place of particular elements of other
> elementary strings in the sentence. [Thus,] there is
> no problem with discontinuous elements in a string
> grammar : all elements which depend in some way on each
> other grammatically appear in the same string or in
> strings which are contiguous by adjunction. (Salkof and
> Sager 1967, 1 and 12).

Roughly speaking, a center string is the skeleton of a
sentence and the adjuncts are modifiers. An example
of a left adjunct of N is the adjective <u>green</u> in <u>the green
blackboard</u>. A right adjunct of N is the clause <u>whom we
met</u> in <u>the man whom we met</u>. A replacement string of N is,
for instance, <u>what he said</u> in the sentence <u>what he said
was interesting</u>. The same sentence with a noun instead of
a noun-replacement string might be <u>The lecture was
interesting</u>. Examples of sentence adjuncts are : <u>in
general, at this time, since he left</u>. The C-strings (...)
have coordinating conjunctions at their head. An example
is <u>but left</u> in <u>He was here but left</u>. Examples of center
strings are <u>He understood</u> and also <u>We wondered whether
he understood</u>. (Sager 1967, 158).

The limitations which a given language imposes on
sequences of the main <u>classes</u> or parts-of-speech categories
can apparently be stated by means of a fairly limited number
of axiomatic formulas : the first published version of a
string grammar of English, in which Harris was mainly concerned
with such general classes, contained no more than eleven
pages of formulas (Harris 1965, 29-40). In writing a more
refined grammar, however, one is faced with the problem of
expressing limitations on the sequences of <u>subclasses</u> :

For example, in the sequence QN1PN2 (Q a number ; P,
preposition), e.g. <u>five feet in length</u>, N1 and N2
are of particular subclasses: x<u>five feet in beauty</u>.
(Salkof and Sager 1967, 1).

Salkof and Sager apply the term 'restrictions' to the
limitations to be stated on sequences or dependencies of
general classes, and the term 'restrictionless string
grammar' to a string grammar in which dependencies of
<u>subclasses</u> are stated by increasing the number of categories
and of string formulas : e.g. N1, etc., instead of N, and,
instead of AN for all Adjective Noun strings, A1, N1, etc.
One most obvious use of distinguishing subclasses i to
predict ambiguities, and to assign the correct structure(s)
to sentences. For instance, they will serve to establish
the difference between the PN strings <u>with interest</u> (analyzed
as a sentence adjunct in <u>he opened the box with interest</u>)
and <u>with a blue cover</u>, a possible adjunct of <u>the box</u> for the

save center string. More fundamentally, a restrictionless
string grammar can provide for a given sublanguage a surface
grammar th t meets the __distinctiveness__ requirements discussed
above, and it can do this without obscuring or "masking" the
relationships between the strings, unlike what Sager seems
to suggest (Sager 1967, 162) ; on the contrary, the subclasses
and the strings in which they occur are much more explicitly
related to each other in such a string grammar than in a
more 'general' one such as Harris's first version. In fact,
a recent report on unpublished current research by Sager
and co-workers of the New York University String Project
confirms the general conception put forward in Harris 1960 and
exemplified below in a passage on biochemical sentences
(Harris 1968, 153) that the grammar of a given sublanguage
"intersects" the grammar of the whole of the language but is
to be regarded as having distinctive properties, i.e. proper-
ties not shared by other subgrammars, or by the grammar of
the language as a whole :

> Recently it has been shown that the language of a
> natural science uses a grammar peculiar to itself. This
> particular grammar is not a subgrammar of the main
> grammar. It brings out word classes and sentence rela-
> tions that are peculiar to the science which the articles
> treat. It can be used for a critique of the formulations
> made in scientific formulations of the concepts and
> relations of that science. (...) The N. Y. U. String Groups
> is about to begin work on a science sublanguage grammar
> within the field of pharmacology (__ACM-SIGLASH Newsletter__,
> 2, 5, December 1969).

As this recent report suggests, the writing and
computerization of a sublanguage string grammar by Sager and
co-workers (and, as far as I know by any other of Harris's
disciples) is still a matter of on-going research, as is the
case for the establishment of criteria for setting up the
subclasses and the string formulas combining them in such a
grammar. I cannot help returning here to my basic point :
one wonders how Sager's team, which I presume consists of
linguists and computer scientists, will establish a the

basis of a __linguistic__ theory what concepts are needed by
experts in the field of __pharmacology__. Other extensions of
Harris's original string grammar are even more problematic,
whether this be after Harris's transformational proposals
(as Sager seems to contemplate : Coyaud 1967, 27) or after
both Harris's and Chomsky's transformations (Shapiro 1967),
or along other lines.

In the present study, I had to face both problems -
that of writing a sublanguage string grammar and deciding
on criteria for doing so, and that of finding approaches to
an extension of such a grammar - together with the additional
problem of discourse or text analysis. In order to outline
the approach which this study embodies for setting up a
variety of a sublanguage string grammar, and for dealing
with relevant problems of discourse structure, I wish to
explain briefly how and why I have adapted Harris's original
proposals on string grammar.

For one thing, the grammar presented below is of the
"restrictionless" variety. It operates on subclasses rather
than on general classes. What is retained from Harris's
string theory is the general assumption about the possibility
of representing the acceptable sentences of a language or
sublanguage as strings of contiguous categories by means of
lists of string formulas of the type illustrated above.
As will be noted, this conception has obvious advantages in
computational analysis.

Secondly, the word subclasses, and consequently the
string formulas which serve to represent sentence parts and
sentences as a whole in the system described in Part II
of this thesis have been established on the basis of the
__conceptual__ criteria yielded by the classification system used
as a __metalanguage__. As a result, the string grammar presented

in this thesis is more akin to Harris's latest proposals
(Harris 1968, which was not available to me at the time)
than to his proposals in the first version of string analysis
(Harris 1965). What I propose, then, is a type of represen-
tation similar to that illustrated in a passage of Harris
1962 (152-153), in which he proposes to represent
a sentence <u>The proteins were treated with acid</u> as
$\underline{N}_{mol}$ <u>is</u> $\underline{V}_{sol}$ <u>en</u> <u>in</u> $\underline{N}_{sol}$. The difference, however, is
that in the present study the subclasses are set up on the
basis of the fact that a given sentence, or rather the
abstract in which it occurs, is relatable to a classification
entry - whereas Harris does not say how he establishes
subscripts such as these used in the above example.

The major shortcoming of my attempt is the following :
while my establishment of the subclasses is determined by
the natural language, I have been unable to be explicit about
how the met language determines the nature of the subclasses
I set up. These, in other words, are unfortunately
open to the criticism I have addressed in 1.1.4. to Katz's
semantic markers : my lexical subclasses belong to a 'private'
metalanguage of mine. The only way out of this difficulty is,
I believe, to be found in the deep structure considerations
presented in Part III of this thesis and directed towards the
establishment of the kind of "syntactico-lexicologico-
semantic" theory now under development by Fillmore and others.

In addition, because of the central role played by the
conceptual metalanguage, the string grammar presented in this
thesis abandons Harris's conception of "center strings" as
the core or central part of sentence representations, together
with the underlying, rather unclear, concept of "sentence-
hood". As Coyaud has noted (Coyaud 1967, ??), there are
many examples of center strings that are not central as
far as semantic content is concerned, and, in this thesis,
I assume that this fact can only be explained in semantic

deep structure. In such cases, it seems that the basic dis-
tinction posited in string analysis between the center
string (and its constituents) and what is adjoined to it
or to the constituents (adjuncts of N', of V, etc.) is
semantically irrelevant. In fact, the adjunction relation
even seems to obscure semantically relevant relations in the
sentence : a semantic _unit_ will often consist of _both_ the
adjuncts _and_ what they are adjoined to. Consider the following
sentence from one of the abstracts under scrutiny in this
thesis : The system described employs three established
techniques : micro-filming, punched-cards, and xerography
(Jansko 1962, abstract 35). If, after Harris (1965,
respectively 23, 29, 30), the center string is defined as
what remains when "no further excisions are possible preserv-
ing sentencehood" and the possibility is allowed of subject
and object Nouns occurring in the center string with such
categories as article, the center string obtained for the
above sentence will be The system employs three techniques,
and all the other words will be analyzed as adjuncts. This
center string is undoubtedly an 'acceptable' sentence. But
what is distinctive about the meaning of the original sentence
is lost with the excision of the adjuncts, and it seems that
this distinctive meaning can only be captured by considering
adjuncts and center string _together_. In fairness to Harris,
I must point out that in Harris 1968 a far more interesting
conception of acceptability is presented than the 'black-
and-white' one which underlies the original version of string
analysis. In Harris 1968, acceptability is a matter of _degree_,
and this conception is central to the whole book, in particular
to Harris's transformational apparatus. Such a conception,
however, is nothing but a proposed parallel to the conceptions
of _deep structure_ advanced by and after Chomsky, and in this
respect, I feel that Harris 1968 has no proposals that would
be equivalent to those discussed below under the heading
'Structural semantics', in the sense that - however elaborate
it may be - his apparatus of transformations has not captured,

or expresses _indiscriminately_, notions discussed in Part III
such as "presupposition", "case", and "implication". His
proposal about the reconstruction of metalinguistic conjoined
sentences (CS ; remember the "Prince of Cambodia" example :
Harris 1968, 133) somehow intersects Fillmore's presupposition,
as well as "implication". His casual mention of "the boy as
the actor ("subject" of _say_)", for instance (_ibid._, 210),
shows his awareness of the "case" problem. And his proposals
for "transformational" derivations of lexical items from _X is_
classifier sentences are relatable to Lyons's implication :
e.g. the derivation of a sentence with "_a large salmon_"
(or rather a discourse consisting of this sentence and one
following, with a 'coreferential' item "_The fish_") from a
pair of conjoined metalinguistic sentences including "_a salmon_
is a fish" (ibid., 167).

The third type of adaptation of string analysis which
has been attempted in this thesis consists of treating texts
(the abstracts under investigation), including their titles,
as conjoined sentences. The hypothesis underlying this
attempt has been put forward by Katz and Fodor, though to
my knowledge never explored by these or by other authors.
In addition, I have extended Katz and Fodor's category of
"sentential connectives" to include the period as well as
other similar pauses of written discourse. Here again,
Harris's well-known claims about the 'grammatical' nature,
and about the neutrality of his discourse analysis operations
towards semantic criteria is my reason for referring to Katz
nd Fodor rather than to Harris :

> In the great majority of cases, the sentence break in discourse is simply and-conjunction. (In others, it is but, for, or, and so on). Hence, for every discourse, there is a single sentence which consists of the sequence of n - sentences that comprises the discourse connected by the appropriate sentential connectives and which exhibits the same semantic relations exhibited in the discourse. (Katz and Fodor 1964, 490-1)

To sum up, the conception of surface structure delineated here is based on Sager and co-workers' published proposals for a restrictionless string grammar, though with two important qualifications. Harris's and his co-workers' conception is used only a means of handling the string aspects of discourse (i.e. the manifestation of discourse as a sequence of contiguous words) by setting up ordered sets of rules, operating on strings of 'grammatical' subcategories, and amenable to mechanization. Secondly, the reduction of the discourse strings leads to distinctive representations of the abstracts, which are represented by entries of the documentary concordance used as a metalanguage. Lastly, no attempt has been made to relate the surface structure representations thus set up to deep structures, because I found that an attempt along these lines is beyond the powers of current linguistic theories, at least as far as the problems discussed in this thesis are concerned. In other words, this thesis does not attempt to formulate transformational rules.

1.3.3. <u>Structural semantics and deep structure representations based on logic.</u>

After Lyons 1963, a structural approach to semantics can be defined as one in which linguistic meaning is not defined as such - in terms of (mostly controversial) psychological or philosophical conceptions about the 'nature' of meaning such as those mentioned in 1.1.5. but is defined instead in terms of empirically discoverable and verifiable facts, namely :

> as a function of (...) several relations (...). In other words, a is not synonymous with b because of its meaning; the fact of their synonymy is part of their meaning. (Lyons 1963, 58)

As Lyons has suggested, it is difficult to imagine how the meaning of such words as <u>truth</u>, <u>beauty</u> and <u>goodness</u> (or, if you wish, <u>cancer</u>, etc.) can be stated otherwise than within such an approach (Lyons 1968, 401). And, if this can be shown, the same will be true of any other word. Underlying Lyons's formulation is the logicians' principle of abstraction :

> What is the weight of a body ? It is usually conceived as an abstract property of the body, recognizable from certain physical effects. Using Russell's principle of abstraction we can reduce the concept <u>weight</u> to the relation <u>having the same weight</u>. The weight of a body is the class of all objects having the same weight as this body. An adept in traditional logic would object that in order to define the same weight we must first define the weight, and then proceed by addition of the differentia specifica to the genus. But there is no reason to insist on this impractical method. It is admissible to conceive the notion of the same weight as prior to that of weight and to define the latter in terms of the former. This conception corresponds to the actual procedure used in the empirical ascertainment of the weight of a body. The balance is a device which indicates, not the weight, but equality of weight. (Reichenbach 1947, 210, quoted in Lyons 1963, 58-9)

One of the most important recent developments in structural semantics has resulted from attempts made by several linguists and other students of language to extend and adapt "predicate calculus" or, perhaps more aptly, the calculus of propositional functions which logicians have set up, so that it can serve the purposes of semantic representation. Here, a few remarks about the linguists' terminology are necessary. Since the emphasis of the studies under discussion is on _extensions_ of the logicians' framework, some terminological usages have resulted to which logicians would probably object, but which I suggest to regard as innocuous in linguistics - as long as the linguist is aware of them : for instance, the use made of such words as "sentence" and "proposition" for what would strictly speaking be regarded as a "propositional function" in the logicians' jargon.

> Going by some of the handbooks of logic with which I am acquainted (in particular Chauvineau 1962, and Gochet 1970b)"functional logic" or "the logic of propositional functions" is concerned with expressions (propositional functions) such as _x is red_, _x is the brother of y_, _x gives y to z_, or with "functions of functions" such as _X is more difficult than Y_ (where X and Y stand for more elementary functions); the logician is particularly concerned with what in the form of these expressions is relevant to their truth-value. Functions are said to consist of "constants" or "predicates" such as _is red_, etc., and of one or more "terms" ("variables","arguments" : X, Y, Z, etc.). These can be specified (replaced by constants), and modified in various ways : by means of "quantifiers" (corresponding to natural language words such as _some_, _all_, _no_) or of more complex expressions (_there is some x such that_).

The adaptation of functional logic in TG linguistics
has given rise to two main theories. One has been initiated
by Fillmore, whose proposals will serve as a basis for Part III
of this thesis, next to recent work on conjunction (in
particular R. Lakoff 1970). The other trend is known as
"generative semantics". This new form of TG theory has been
developed by and around such linguists as George Lakoff,
James McCawley, and Paul Postal. I shall not discuss "gene-
rative semantics" in this thesis or attempt to compare the
new theory with Fillmore's, since such a comparative study
would duplicate a recent paper of mine (J. Noël 1971)
and would not be pursued in Part III, which is based on
the rival theory proposed by Fillmore. Still, it is important
to realize that the problems of semantic representation
discussed in Part III could presumably have been approached
on the basis of generative semantics, and that, if such
an exploration of my problems on the basis of this theory
were available, it would be possible to carry out an
evaluative study of the two approaches - a task which, to
my knowledge, has so far not been attempted systematically.

1.4. PROCEDURE

1.4.1. Towards a linguistic study of 'pragmatic' competence.

The question raised here is in what sense the semanticist's task can be described broadly as that of explaining "what underlies the use of the word means and related terms (...) in English and other languages" (Leech 1969, 5). In brief, I shall argue in favour of a linguistic study of 'pragmatic' competence, which in this study has its concrete counterpart in the proposal to consider metalanguage, theory and procedure as the main components of semantic analysis. Negatively, I shall argue that a notion of pragmatic competence should give up the following main features of Chomsky's notion of competence:

(i) the indeterminacy, discussed in 1.1.3., of Chomsky's metatheoretical conception of the boundaries separating the domains of linguistic and "pragmatic" competence, and of syntax, semantics, and "systems of knowledge and belief";

(ii) Chomsky's and his followers' commitments, mentioned in connection with the linguist's neutrality (1.1.5.), to various psychologically and philosophically controversial issues : in particular, Chomsky's negative attitude to communication, Katz and his associates' speculations about the universality of semantics, and Chomsky's mentalistic speculations on intrinsic or innate competence.

> Most recently, Chomsky (1969a, 83) put this as follows :
> "We may formulate the problem as the problem of determining
> the intrinsic characteristics of a device of unknown proper-
> ties that accepts as 'input' the kind of data available to
> the child learning his first language, and produces as
> 'output' the generative grammar of that language. The 'out-
> put', in this case, is the internally represented grammar,
> mastery of which constitutes knowledge of the language."

(iii) the undue importance given by Chomsky to the difference between competence - establishable only on the basis of the linguist's introspection - and performance, and his misleading equation of other observable data with performance viewed as a range of facts of secondary importance for theoretical

investigation, on the ground that competence in observable data is irretrievably distorted by such factors as "memory limitations, distractions, shifts of attention and interest, and errors " (Chomsky 1965, 3).

(iv) the _practical_ equation (in his analyses) of "sentences" with theoretical objects ALWAYS yielding surface structures of sentences that are close to sentences in the everyday sense of the word, and his lack of concern with structures that would correspond to discourses (see the unexplored suggestion of Katz and Fodor).

Positively, however, I propose:

(i) to adopt Chomsky's conception of the linguist-observer's observations as facts of performance;

(ii) to recognize the existence of a 'gap', as it were (cp. Leech's expression "what underlies"), between the linguist-observer's performance and linguistic competence, a gap which Chomsky discusses in terms of the "creative" aspects of language, and which I have found useful to discuss below around a notion of "inaccessibility";

(iii) to define the "ideal speaker-hearer's linguistic competence" roughly as made up _several competences_, describable by the following formula : _conceptual expertise statable by a metalanguage plus facts statable by linguistic theory equals 'competence'_;

(iv) to consider that competence thus defined is ideally describable by means of "_precise and complete procedures_" (Gardin), and that this goal, like any other scientific undertaking, is a sisyphean one (Gardin, personal communication).

In what follows, then, I would only agree with Chomsky's emphasis on "appropriateness", on the "creative" aspects of language, and on his consideration of the linguist-observer's observations as facts of "performance" serving as a basis for an account of competence as described above.

In brief, Chomsky has defined the task of linguistic theory
as that of setting up rules accounting for an ideal hearer-
speaker's competence, or "perfect" knowledge of the language.
This is defined as an ability to produce and understand an
indefinite number of "appropriate" sentences, including
sentences never heard or spoken (or read) before. This
competence is distinguished from performance, the way in
which people actually use the language. The assumption is
that performance can be defined at a later stage of research,
in terms of departures from or differences with theoretically
defined competence :

> Linguistic theory is concerned primarily with an ideal
> speaker-listener, in a completely homogeneous speech-
> community, who knows its language perfectly and is
> unaffected by such grammatically irrelevant conditions
> as memory limitations, distractions, shifts of attention
> and interest, and errors (random or characteristic) in
> applying his knowledge of the language in actual
> performance. (Chomsky 1965, 3)
> We thus make a fundamental distinction between competence
> (the speaker-hearer's knowledge of his language) and
> performance (the actual use of language in concrete
> situations).(ibid., 4)

One of Chomsky's main concerns is how linguistic theory
can provide "an explicit formulation of the "creative"
processes of language", "a real understanding of how a
language can (...) "make infinite use of finite means""
(ibid., 8), particularly of how language provides the means
for "reacting appropriately in an indefinite range of new
situations". One of the basic claims of this thesis is that
the problem of the indefinite number of METAlanguages is one of
the essential problems involved :

> an essential property of language is that it provides
> the means for expressing indefinitely many thoughts
> and for reacting appropriately in an indefinite range
> of new situations. (...) The grammar of a particular
> language, then, is to be supplemented by a universal
> grammar that accommodates the creative aspect of language
> and expresses the deep-seated regularities which, being
> universal, are omitted from the grammar itself.
> (ibid., 6)

Linguistic observations must be regarded as "data of performance", and :

> The problem for the linguist, as well as for the child learning the language, is to determine from the data of performance the underlying system of rules that has been mastered by the speaker-hearer and that he puts to use in actual performance. (ibid., 4)

To sum up, the purpose of this study is to exemplify and explore the suggestion that metatheoretical recognition needs to be given to the non-linguistic or conceptual foundations of linguistic competence, in the form of a metalanguage (or in some other form), and that "appropriateness", which I believe to be a central feature of competence, is theoretically statable by giving a primary place to conceptual facts in semantic studies. If it is to avoid being "trivial" and to drop "pretensions of omniscience" (Bolinger 1965), a conception of "creativity" and of the novelty of linguistic utterances - the other main feature of competence - will similarly have to recognize the conceptual foundations of linguistic competence. Among other things, it will have to account for creativity in science and literature : for instance, for the circumstances under which people can speak of machine translation (whatever the corresponding 'thing' or reference may be), or for the impact of people like Shakespeare and Chomsky on languages and sublanguages. While Chomsky's theory of syntax has stated and explored mathematical properties of language that account for creativity, I am not aware of the existence of equivalent proposals that would be relevant to the epistemological problems of people's conceptual or semantic creativity. Under these circumstances, I feel that the best thing to be done in these prolegomena is to describe the problem rather than attempt to offer solutions.

After the metatheoretical proposals outlined above and after
Chomsky's conception of linguistic observations as data
of performance, I shall go out from the communication situa-
tion which I propose to investigate in this study. The problem,
as I wish to present it, concerns the inaccessibility of
both object-language and metalanguage (though in different
respects). The competence to be defined is the documentalist-
indexer's or EXPERT-indexer's ability to represent the meaning
of an indeterminate number of linguistically undefined abstracts
in his field by means of one among other finite and linguistical-
ly definable metalanguages for this field. The problem of
'semantic' creativity is THE problem, but can hardly be touched
upon, for lack of reliable knowledge in this area. Perhaps
a solution will come from disciplines other than linguistics.

As I deputize for the expert in this study, my unachiev-
able goal as an observer is a definition of competence on the
basis of my observations of "only an inconsiderable subset"
(Leech 1968, 94) of all possible abstracts on documentation.
The object-language is 'inaccessible' in this sense. The
inaccessibility problem, as it concerns both the object-language
and the metalanguage, can be described by asking, with Stone
(et al., 1966) :

> How can the infinite variety of words, phrases, sentences,
> and styles be the source of stable, scientific conclu-
> sions ? (ibid., 4)
> How does the investigator know when he has specified
> all the alternative ways in which a characteristic can
> be expressed ? He usually does not. On making further
> applications of his categories, he will often find new
> and unanticipated alternatives. He may fail to remember
> that certain words and phrases are signs of a social or
> psychological variable. He may also forget about the
> usages of a word or phrase under certain circumstances,
> which differ from those he has in mind. (ibid., 8)

Of course, "most of these errors can be avoided if the investigator collects a number of instances of actual word usage in context" (ibid.), but to do this only is to accept defeat, as the purpose of a theory is precisely to offer substitutes for data collection, and to predict what otherwise could only be obtained by observation of texts.

The inaccessibility problem with respect to the meta-language is somewhat different. If we do not assume that the metalanguage consists, by definition, of a finite number of terms and relations and that these are available prior to the investigation, the inaccessibility problem is that of stating when, why and how the metalanguage will need to be changed. Assuming that the metalanguage is finite and fixed once and for all, as I have done in this study, there is still an inaccessibility problem in that we have no guarantee that the formulations of the metalanguage are the only, or the best possible ones, even with respect to a single 'class' of observers. Since the metalanguage used in the present study is a kind of classification consisting of definitions, the following passage from Langendoen on the question of dictionary definitions most appropriately illustrates the problem of inaccessibility as it arises here for the metalanguage.
It will be noted that Langendoen does not capture, or does not wish to use the notion of metalanguage. This notion, however, is obviously implicit in his reference to a "prescriptive (...)" standard language", and in the statement he makes that "sentences only take on meaning upon conscious reflection", in the passage quoted above (in 1.1. 2.) about "legal documents" belonging to langue :

> The experience of lexicographers (...) is that meanings
> for lexical items are extraordinarily difficult to pin
> down, even if dialect differences are overlooked (we may
> view the task of a lexicographer, at the risk of sounding
> prescriptive, as dealing with the meaning of an item
> in a standard language); one can almost never be sure
> that some element has been overlooked, or conversely that
> some superfluous material has been included. Alternative
> ways of defining items are almost always possible, and
> considerations of simplicity or elegance are often of no
> help in choosing among alternatives, and may be downright
> misleading. (Langendoen 1967, 107)

On several occasions, Gardin and co-workers have made
similar observations about the metalanguages used in documenta-
tion. They point to the frequent difficulty of choosing between
several possible relations among those defined in SYNTOL, the
model which Gardin has set up for semantic representation
in automatic documentation (Cros et al. 1964, 45-6). The
authors of SYNTOL insist that, in many cases, none of the
eligible relations can be said to be 'better' or more appro-
priate than any other. Similarly, it is difficult to say
which metalanguage symbol is best for naming a given
concept. In some cases even, a given natural language
has no word corresponding to the metalanguage concept needed
for documentary analysis : cp. "the absence of a word that
bears to plants the relation that "corpse" bears to animals"
(Chomsky 1965, 232).

The relevance of what has been said so far to the present
thesis is most obvious in the many possibilities that exist
of naming the very field to which the abstracts under investi-
gation belong : (scientific and technical) information, documenta-
tion, information storage and retrieval, retrieval system theory,
information science (and technology), scientific communication,
knowledge availability, etc. While these expressions and
many others (including various lengthy paraphrases and
definitions) are often reducible to a single metalanguage unit,

It is difficult to decide which expression best names the
unit in question and, until the analyses presented in this
thesis can throw some light on the problem, the only thing
to do is to make an arbitrary choice, for instance in favour
of "documentation", as I have done in what precedes. One of
the crucial theoretical problems, under the conception so far
described, is that of defining "dispersion factors" (Gardin
1969) that will account for one of the main symbolic aspects
of people's 'metalinguistic' activities. It is the fact that
a given metalinguistic unit or structure is broken up into
formally different or more complex units or structures in the
object-language, and vice-versa. It is a fact most dramatically
observable in the traditional library practice of reducing
a whole book to a handful of symbols drawn from a classifica-
tion, and of using such a classification as a means of access
to the knowledge shared by the members of our communities -
and, indirectly, of extending this knowledge creatively.

1.4.2. From a linguist-observer's performance to competence.

I can now outline the ideal objective of this thesis as that of converting "observational" statements into an integrated set of other types of 'public' statements, in order to be able to describe the achievements and shortcomings of the present study with reference to such an ideal. This ideal, as far as the topic of this thesis is concerned, is the setting up of a "precise and complete" procedure based on a perfect linguistic theory, and relating a particular object-language (the English of abstracts on documentation) to one of the corresponding metalanguages (a classification concordance for the field of documentation) by means of "explicit" rules.

The empirically prior statements are "observational" statements (Leech) of the type illustrated in Figure 2. They express judgments of mine (for lack of approval or dis-approval by competent informants) about the relations between texts of the object-language under investigation and the metalanguage used. This metalanguage is one of the possible definitions of the conceptual field of documentation. Its use, or the use of another for the same field, is indispensable in this study since linguistic theories are unable to define this (or any other) conceptual field and such a definition - as I have argued - needs to be made explicit as soon as one wishes to engage in a semantic analysis.

A brief actual example may be appropriate here to suggest what problems arise from the observational statements I go out from, and to illustrate the distinctive function of these observational statements. The example distorts my point only insofar as it can not illustrate the inaccessibility problems discussed in the preceding section : for a perfect illustration in this respect, one would

need to consider all the entries of the classification used, and a perfect formulation of these entries and of their interrelations on the one hand, and the total possible set of abstracts in the field of documentation, on the other. Perhaps one would also need to consider what non-documentalist observers make of abstracts on documentation, however speculative the task may be. Here then, my example will consist only of two simplified classification entries - (1a) and (2a) respectively - and two corresponding abstracts (one of which is quoted only in part), (1b) and (2b) respectively. Thus, the observational statements which serve me as a starting point are illustrated by the relation I have established between (1a) and (1b) on the one hand, and (2a) and (2b), on the other. In other words, the a expressions are metalinguistic, the b expressions belong to the object-language, and the problem is how the relations established empirically between a and b can ideally be expressed in the form of rules.

(1a) Scientific and technical information institutions described individually.

(1b) The Japan Information Center of Science and Techno-logy and its activities (title).
A general description of the Japan Information Center of Science and Technology which was established for the purpose of contributing to the development of science and technology by collecting, processing, storing, and retrieving scientific information (sentence 1 of abstract 1 : Janaske).

(2a) General organization of scientific and technical information.

(2b) Report on a study of scientific communications for the National Science Foundation (title).
This reviews the initial reasoning and thinking behind the design and construction of a project supported by the National Science Foundation, carried out by the Operations Research Group at the Case Institute of Technology to determine ways "to affect the behavior of scientific institutions that are responsible for the dissemination of recorded information in order to achieve a net improvement in scientific productivity?"(text of abstract 2, consisting of a single sentence : Janaske).

PART II : AN EXPERIMENT IN MECHANIZED INDEXING.

2.1. INTRODUCTION.

2.1.1. Background and objectives.

The mechanized indexing system presented here was
developed at the University of Victoria between May 1968
and May 1969. Preliminary tests, which were run between
September 1967 and May 1968, will not be reported here,
since they are described elsewhere (my unpublished M.A.
thesis : J. Noël 1968) and have been superseded by the present
system. This system has analyzed the first 50 abstracts of
the Janaske bibliography (Janaske 1962). The edited version
of these 50 texts as they were actually processed by the
computer is appended (Appendix 3). The unedited version of
the first 5 abstracts of the Janaske collection is quoted
in Appendix 4, that is, the version in which they actually
appear in Janaske with bibliographic information which was
not punched on cards and with various other types of informa-
tion which were edited out before and during the punching
(see 2.2.2.). The main entries ('groupes') of the Gardin-
Lévy concordance which serves as a metalanguage in this
study are quoted in Appendix 1. My own English translation
of the main headings of this list of descriptors is given
in Appendix 2.

Henceforth, I shall abbreviate my references to the
Janaske bibliography as follows : (J 1) will mean '(quoted
from) abstract 1 of the Janaske corpus', etc.

The computer programmes written for the mechanized
indexing system will not be described here. They were written
in PL-1 for the IBM 360/44 computer of the University of
Victoria. The programmer, Rémi Michelot, has so far not
published his programmes. Here, I shall therefore only
report on my own work, which consisted of setting up a system
that could be programmed during the time available, on the
basis of the metatheoretical assumptions presented in Part 1
and of considerations to be discussed presently.

First, I must point out that the system presented here
has not been set up in a real retrieval situation (on this,
Salton 1968). In particular, I have not been able to have
my indexing (2.2.1.) checked by users in an actual retrieval
environment.

Furthermore, no computerized syntactic analyzer or
indexing system has been available to me during or after
the setting up of the system. Nor have I been able to make
use of a ready-made computerized dictionary of documentation
English; in fact, I do not know of the existence of such a
dictionary in the literature. Under these conditions   , I
decided to set up my own system, and to do this on the basis
of various preliminary simplifications and assumptions.

The main preliminary simplifications which were decided
are the following :
(1) The experimental corpus was limited to fifty abstracts
    in their edited version as given in Appendix 3.

(11) As descriptors, I used my English translation of the
main headings of the Gardin-Lévy concordance (Appendix
2), and these descriptors were left unanalyzed and
treated as a single symbol. As a result, it was possible
to reduce the problem of indexing to that of representing
each abstract of the experimental corpus by one descriptor
(only occasionally more than one) treated as a single
symbol, and to reduce the problem of mechanized indexing
to that of 'translating' each abstract into a single
code (see 2.4.9.) standing for this (these) descriptor(s).

(111) Many automatic syntactic analyzers and indexing systems
are set up in such a way that in a given situation the
machine can choose between a number of decisions, some
of which can be judged to be wrong by informants. Thus,
all known syntactic analyzers produce several analyses
for a given sentence, and only some of these analyses
can be described as 'syntactically' and/or 'semantically'
'correct'. Similarly, Bély et al. ,who have published
the largest and most sophisticated automatic indexing
experiment known to me, have included in their final
report an evaluation study in which various "correct"
and "incorrect" computer outputs are discussed (Bély et al.
1970, Ch. 6 and Ch. 12). In the experiment presented
below, I have adopted a different approach. From a compu-
tational point of view, the purpose of the experiment
was to mechanize a system which was previously set up
and tested manually and which, from the outset, was
assumed to produce the 'correct' output. As it appeared
from the manual tests and from consultations with computer
experts that the system worked and was programmable, the
programmes were written and the system was run success-
fully from beginning to end on the computer. As a
computational experiment then, the system is a complete

success, thanks to the devotion and (for me, most instruc-
tive) cooperation of the programmer. The only qualifica-
tion is that, within the year available for the experiment,
it was not possible to integrate all the operations
performed by the system into a single programme. Thus,
in the system as it stands, human intervention is necessary
to use the tape with the output of one stage of the system
as input to the next stage.

(iv) For practical reasons, it was not possible to test the
system with abstracts outside the experimental corpus,
for instance some of the remaining 1121 abstracts of the
Janaske corpus.

As for the <u>linguistic</u> aspects of mechanized indexing,
the system under discussion was set up to test an approach
to mechanized indexing which can be described as follows :
(1) The sentences of a text can be represented for computer
processing <u>as grammatically defined strings of CONTIGUOUS
word classes</u>, and a text can in turn be represented <u>as a
string of contiguous sentence strings</u>. This is a conception
of language most clearly formulated by Harris. More
broadly, the assumption of contiguity can be regarded as
a basic postulate of linguistics, <u>and</u> as a prerequisite
for assuming that language data are computable. If,
for instance, a rule states that a string consists of
<u>A</u> and <u>B</u>, no other symbol is allowed between <u>A</u> and <u>B</u>,
and the computer is programmed to recognize just the
string <u>AB</u>. In one of Harris's recent formulations :

> Talk or writing is not carried out with respect to
> some measured space. The only distance between any
> two words of a sentence is the sequence of other
> words between them (...) Hence, the only elementary
> relation between two words in a word sequence is that
> of being next neighbors (...) Any well-formedness
> for sentence structures must therefore require a
> contiguous sequence of objects, the only property that

> makes this sequence a format of the grammar being
> that the objects are not arbitrary words — words of
> particular classes (...) But the sequence — s to be
> contiguous; it cannot be spread out with spaces in
> between, because there is no way of identifying or
> measuring the spaces (Harris 1968, 16).

The "elementary relation" referred to by Harris is
concatenation, and will be symbolized as + , as in '$\underline{A+B}$'
standing for a string $\underline{AB}$. '$\underline{A+B}$' means that $\underline{A}$ is
immediately followed by $\underline{B}$.

(ii) From (1), it follows that, by means of concatenation
rules, a text can be represented as a single string,
consisting of smaller strings. The concatenation rules
are based on a grammar : "the only property" that makes
the sequence of words in sentences and in texts "a format
of the grammar" is that the words are "not arbitrary
words but words of particular classes" (Harris, loc. cit.).
Because the grammatical concatenation of word classes is
the only fact about the surface structure of language
that can be assumed to be true of any language variety,
it was found that the mechanized indexing system presented
below would achieve greatest generality by being based
essentially on this fact. Moreover, even under a conception
of language which distinguishes between semantic deep
structure and grammatical surface structure, a treatment
of the concatenation facts of surface structure is a
necessary (though not sufficient) condition for an
analysis of language data.

(iii) In agreement with the requirement of distinctiveness
discussed in Part 1, the concatenation rules are stated
for "subclasses" (such as N_{mol} for "proteins" in
biochemical language : Harris 1968, 153) rather than for
classes (such as $\underline{N}$ for Noun) holding for the language
as a whole. This raises the problem of subcategorization.
Irrespective of this problem, the string grammar set up

for the system under discussion is of the "restrictionless" type, that is, one which states dependencies of __subclasses__ - with a resulting increase in the number of categories and of string formulas (Salkof and Sager 1967).

(iv) In the system, the problem of subcategorization is separated from that of setting up concatenation rules. In other words, the problem of subcategorization is treated as one of dictionary-making which is independent of the system proper. In this way, the dictionary codes can be changed or improved without affecting the general organization of the system. For obvious computational reasons, however, a __format__ for lexical codes had to be fixed in terms of the maximum number of characters which lexical entries can consist of. I believe that there are good reasons for adopting this treatment of subclasses. I feel that lexical problems in general, and problems of subcategorization in particular, are the most poorly known area of linguistic theory. Besides, I have not succeeded in setting up a satisfactory approach to sub-categorization for my mechanized indexing system, as will appear from the discussion of subcategorization presented below (2.3.).

(v) No known theory is able to predict the length - either minimal or maximal - of the strings that serve to express a given meaning. In particular, as far as the present study is concerned, there is no way predicting or deciding __a priori__ which strings of the abstracts under investigation will correspond to a given entry or descriptor of the meta-language. In the system, this matter has therefore been treated as a descriptive problem, and has been left open to further research. To take just one well-known example of the difficulty alluded to here, consider abstract J 36 (Appendix 3), whose last sentence contains a negation

denying some of the information implied by the preceding
sentences (see section 3.3.4. on "implication"). Because
of this difficulty, I decided to set up the system in such
a way that the machine would have to analyze all the words
occurring in each text, thus leaving open the question of
the size of the strings that can be associated with a given
descriptor.

(vi) By virtue of the definition given above of indexing as
the reduction of a text to a single symbol, the various
stages of the system do not only serve to concatenate
strings of grammatical classes. Their function is also
to delete information. The loss of information which
takes place during the process of indexing - particularly
so-called indicative indexing of the type adopted for this
study - is indeed one of the problems which needs to be
faced in any mechanized indexing experiment : obviously,
one or two descriptors are unable to represent all the
meaning of a text. The problem is how to symbolize this
loss of information, in spite of the difficulties just
mentioned in (v). This problem has also been treated as a
descriptive problem in the present experiment, because it
is beyond the predictive power of current theories.

In this thesis, I shall not attempt to improve directly
on the treatment of subcategorization and of semantic units
adopted in my mechanized indexing experiment. I shall assume
that the lexico-semantic problems involved are more fruitfully
dealt with on the basis of a conception of semantic deep
structure than in terms of an extension of string theory. The
essential purpose of Part III of this thesis is to suggest
ways of remedying the semantic shortcomings of the mechanized
indexing experiment on the basis of recent work by generative
linguists on problems of semantic representation.

2.1.2. Synopsis.

In brief, the mechanized indexing system works as
follows. After a dictionary-lookup operation which replaces
each text word or symbol by a code, the machine performs step-
by-step concatenations of these codes, rewriting two or more
codes as a single code by looking up lists of rules stored
in it, until each abstract is rewritten as a single code.
After dictionary-lookup there are in total 10 main sets of
concatenation rules, which apply in succession. Each set of
rules makes use of the results of the preceding set of rules.
The ten sets of concatenation rules are based on only _four_
types of operations performed by four computer programmes,
three of which are used more than once.

The programme which is used only once has the effect
of replacing some proper nouns by a general label for proper
nouns : thus, "the National Science Foundation" is rewritten
as 'NPP' standing for 'proper noun'.

Using letters ($\underline{A}$, $\underline{B}$, $\underline{C}$, etc.) to represent the codes
which the machine concatenates, the operations performed by
the other three programmes can be explained as follows.

The first programme concatenates only two codes at a
time and works from left to right. It can, however, operate
in several passes, each of which makes use of the results
of the pass that precedes. Thus a string $\underline{ABC}$ is processed
in two passes. In a first pass, $\underline{B}$ and $\underline{C}$ are concatenated :
$\underline{B} + \underline{C}$. In a second pass $\underline{A}$ is concatenated with $\underline{B} + \underline{C}$, yielding
$\underline{A} + \underline{B} + \underline{C}$. Alternatively, $\underline{A}$ can be concatenated with $\underline{B}$, and
$\underline{AB}$ with $\underline{C}$.

The second programme concatenates pairs of codes between which there is a code for a <u>conjunction</u> or for some category equated with conjunction. Thus, writing small <u>c</u> for conjunction, this programme concatenates strings of the type <u>AcB</u>, yielding <u>A + B</u>. The rewriting (<u>A + B</u>) obtained by a first concatenation can be concatenated with a following conjoined code. Thus, an enumeration <u>AcBcC</u> can be rewritten in two passes as <u>A + B + C</u>. This programme works from left to right : it begins by testing for the occurrence of a first <u>c</u>, then it tests for the occurrence of an <u>A</u> preceding and of a <u>B</u> following this <u>c</u>.

The third programme concatenates between six and two codes : <u>A + B + C + D + E + F</u>, <u>A + B + C + D + E</u>, <u>A + B + C + D</u>, <u>A + B + C</u>, <u>A + B</u>. The first pass processes strings of six codes, the second strings of five, the third strings of four, the fourth strings of three, and the fifth strings of two codes.

In order to simplify the presentation, I shall now refer to each successive set of concatenations performed by the three programmes just mentioned as a series of 'grammars' : respectively G.0 ('grammar 0'), G.1, G.2, G.3, G.4, G.5, G.6, G.7, and G.8. The following table (1) summarizes the workings of the system, and the various operations it performs. The arrow (-->) stands for 'is rewritten as' (the exact meaning of this rewriting operation in the system will be explained later).

(1)

ORDER OF OPERATIONS: TYPE OF OPERATION:	1	2	3	4	5	6	7	8	9	10	11
Dictionary lookup: text word-->code	X										
Some proper nouns -->NPP		X									
Pass 1:BC-->B+C Pass 2:AB+C-->A+B+C Etc.			C.0				C.4				
Pass 1:AcB-->A+B Pass 2:A+BcC-->A+B+C Etc.				C.1		C.3			C.6		C.8
Pass 1:A...F-->A+...F . . . Pass 5:AB-->A+B					C.2			C.5		C.7	

As appears from this table, the 'conjunction reduction' programme operates after each of the other concatenation operations except C.4. The last application of this programme serves to concatenate the codes obtained for the title and the sentence(s) of each abstract, the code for the period being treated as a conjunction.

On the basis of the conventions adopted in table (1), I can now outline the rest of this presentation. The next chapter is devoted to the work that was done to prepare the system (2.2. PREPARATORY WORK). This chapter is subdivided as follows : Indexing (2.2.1.), Punching and editing (2.2.2.), and System design (2.2.3.). The third chapter is entitled SUBCATEGORIZATION AND DICTIONARY LOOKUP (2.3.).

Chapter 2.4. is a description of the main components of the
system. The first section deals with the analysis of proper
nouns (2.4.0.). The remaining sections (2.4.1. to 2.4.9.)
explain the workings of the various 'grammars' that operate
after the reduction of proper nouns, namely G.0 to G.8. In
the last   chapter , I shall present sample analyses
and draw conclusions (2.5.) concerning what I believe to be
the pros and cons of the system, and possible ways of
improving on it.

Throughout the presentation of the system, I shall try
to concentrate on general problems and principles and to
select significant examples rather than give an exhaustive
description of the system, because such a description would
require several hundreds of pages. As a matter of fact, our
experiment turned out to be a very large one indeed. The
complete computer printout for the system represents two big
volumes of machine lists. It took me a whole week to reduce
these lists to fifty tables giving a synoptic analysis of
each abstract of the experimental corpus (such a synoptic
analysis has not been programmed, as pointed out in the
preceding section), and I have found that even these fifty
tables which give the successive rewritings from G.0 to G.8
are too large and too complex to be presented as such in the
ensuing report. This is the reason why I decided to break
down my account into the various chapters and subdivisions
~~outlined~~ in the preceding paragraph.
mentioned

2.2. PREPARATORY WORK.

2.2.1. Indexing.

As a preliminary step towards the setting up of the
mechanized indexing system, I indexed the 1121 abstracts
of the Janaske bibliography. I felt that this was necessary
for me to gain familiarity with the texts, and with the field of
documentation. Here, however, I shall only report on the
indexing of the 50 texts of the experimental corpus, which
are listed in Appendix 3 in their edited version. My indexing
of the 50 abstracts is tabulated below in (6).

The 25 main headings of the Gardin-Lévy concordance
which were used as descriptors in this study are given in
Appendix 2. Those that were used to index the abstracts of
the experimental corpus - in total 12 descriptors - are
listed below in (2) (a) with the number assigned to them by
N. Gardin and F. Lévy. In (2) (a), STI stands for "scientific
and technical information". Following a suggestion made by
N. Gardin and F. Lévy, I have merged number 9 (Techniques of
reproduction) into number 23 (Machines and equipment).
Abbreviations of the headings of (2) (a) which will be used
in table (6) below are given in (2) (b).

(2)

(a) 1. STI : generalities.
2. Professional problems.
3. Information science
4. Related sciences.
5. General organization of STI.
6. STI institutions.
7. Functioning of STI networks.
8. Sources of information.
10. Information handling : generalities.
16. Classification and indexing.
19. Coding.
23. Machines and equipment.

(b) 1. GEN(eralities).
    2. PROF(ession).
    3. SCI(ence).
    4. SCO(other sciences).
    5. GEN(eral) ORG(anization).
    6. INST(itutions)(= individual institutions).
    7. FUNCT(ioning).
    8. SOURCES.
   10. HAND(ling).
   16. CLASS(ification).
   19. COD(ing).
   23. EQUIP(ment).

Before reporting on the results of my indexing, I wish
to mention some difficulties which I experienced in indexing.

The first difficulty, which has already been mentioned
in 2.1.1. has to do with the size and nature of the strings
of discourse which can be said to correspond to a given
descriptor. As an example, consider abstract J2, which is
quoted below in (3) together with the descriptor assigned
to it.

(3)

(a) Report on a study of scientific com-
munications for the National Science
Foundation.

This reviews the initial reasoning and thinking be-
hind the design and construction of a project sup-
ported by the National Science Foundation, carried
out by the Operations Research Group at the Case
Institute of Technology to determine ways to affect
the behavior of scientific institutions that are
responsible for the dissemination of recorded infor-
mation in order to achieve a net improvement in
scientific productivity. (J2).

(b) <u>Descriptor</u> : General organization of STI.

Let us assume that the text is aptly represented by the
descriptor "General organization of STI" : while the expres-
sion "scientific communications" in the title is ambiguous
(for reasons to be discussed in Part III, particularly in
3.3.4.), the expression "the behavior of scientific institu-
tions that are responsible for the dissemination of recorded
information" is relatable to the descriptor in question, and
to its definition (Appendix 1, "Groupe 5"). What is unclear
(and will be left undiscussed in the rest of this thesis)
is how and why the rest of the text allows for, and does not
preclude, the interpretation in question. In other words,
even though one can identify in the text a passage which
is directly relatable to the descriptor assigned to the
text, there is no known way of determining the contribution
of the rest of the text to the meaning of the 'significant'
passage in question. To return to the simple example given
above, the text would not have been assigned the same
descriptor if the passage about "scientific institutions"
had been the grammatical object of, say, "The author does NOT
discuss——" (rather than of "...determine ways to affect——").

A second difficulty (and another problem left undiscussed
in this thesis) concerns the assumption just made about (3).
As already pointed out, I have not attempted to analyze the
headings of the Gardin-Lévy concordance, that is, I have not
analyzed the French definitions, nor the general headings
of Appendix 1, nor the English translation of these headings
given in Appendix 2, nor the relations between the headings
and the definitions. Obviously, the passage "the behavior
of scientific institutions that are responsible for the
dissemination of recorded information" (J2) (in particular
the plural "institutions") is more directly relatable to
the definition of "Groupe 5" (Appendix 1 in particular

"Ensemble des organismes et institutions qui forment l'appareil
général de l'IST, dans un ou plusieurs pays...") than to the
heading "General organization of STI", and it would be
necessary to analyze the inference process by which we relate
the heading, the definition and the corresponding expression
in the object-language. I have not attempted to do this
because I have found no theory that could throw light on
this inference process : such a theory, for instance, would
have to show the relation between grammatical marks like
plural ("institutions", "organismes") and words like "general",
"ensemble", etc. Another example of the same difficulty is
J32 which has been indexed by the descriptor "STI : generali-
ties". In the definition of this heading, mention is made
of "considerations concerning the future of scientific informa-
tion" (Appendix 1, Groupe 1). The use of the future tense
in the text is one of the clues (next to vague, everyday
adjectives like "fast, large" in 4) for describing its topic
as "generalities", rather than, say, as scientific considera-
tions about "memories" and storage devices (as in the case
in J11, Appendix 3). The last sentence of J32 is quoted below
in (4) :

<div style="text-align: center;">(4)</div>

> Machine translation of languages and recognition of
> spoken information are two other areas which will
> require fast, large memories (J 32).

Part III of this thesis concentrates on problems that are
common to the metalanguage and to the object-language, but
does not clarify the nature of relations that are intuitively
establishable between metalanguage and object-language such
as those just exemplified.

A third difficulty which I have experienced in indexing
is even more serious : I am uncertain about the correctness
of my indexing of some abstracts of the experimental corpus.
These 'difficult' abstracts will be marked below with a
question mark (?). As an example, consider J18 quoted below
in (5) in its unedited version :

(5)

Application of high-speed computers
to information retrieval.

This is a description of the following types of infor-
mation retrieval systems : statistical, syntactical
and interrelational. In addition, the computer research
program of the Patent Office--HAYSTAQ-- is described.
(J18).

The title has a very general meaning, while "<u>statistical,
syntactical and interrelational</u>" in the text suggests that
the topic of the text is narrower : presumably describable
as "Classification and indexing". In such cases, I decided
to assign to the text the broadest possible descriptor, in
this case "Information handling : generalities".

The occurrence of "<u>computer research program</u> of the
Patent Office -- HAYSTAQ" in (5) exemplifies one last
difficulty I wish to mention in connection with indexing.
Unless one has the necessary expertise to know what HAYSTAQ
is, "computer research program" is ambiguous in various ways.
"<u>Program</u>" can mean either "project" or "computer program"
in the sense of routine.Besides "computer research" is
ambiguous between research ABOUT and research WITH(the aid of)
computers. If HAYSTAQ is left out, as has been done in the
edited version of J 18 (Appendix 3; see next section), there
is no way of resolving the ambiguity.

In (6) below, I have tabulated my indexing of the 50
abstracts of the experimental corpus. In order to avoid
duplicating Appendix 3, I merely give passages of the abstracts
in column 1,when there is a short passage - such as a title -
that is rather obviously relatable to the descriptor chosen.
The descriptors are given in the abbreviated form of (2) (b).
The extracts quoted in column 1 should not be regarded as the
'significant' passages of the texts they are drawn from;
as already pointed out, the size of a 'significant' passage
cannot be safely specified or predicted in the current state
of linguistic theories. An attempt will be made, however, in
Part III of this thesis to approach the problem via the notion
of 'topic' developed in Chapter 3.2.

DESCRIPTORS : EXTRACTS WITH Nr of ABSTRACT :	GEN	PROF	SCI	SCO	GENORG	INST	FUNCT	SOURCES	HAND	CLASS	COD	EQUIP
The Japan Information Center...and its activities (J1)						X						
scientific institutions that are responsible for the dissemination of recorded information (J2)					X							
the National Institutes of Health Russian scientific translation program (J3)						X						
problems of a manufacturer ...in searching for information in the medical and chemical literature (J4)							X	X				
The Federal Government and U.S. scientific information (J5)					X							
United States scientific and technical information services (J6)					X							
documentation centers in the German Democratic Republic (J7)					X							
Training the scientific information officer... syllabus...for a... course of training...(J8)		X	X									
the IBM 650 information retrieval system which uses the inverted file (J9)												X

(6) (continued)

DESCRIPTORS : / EXTRACTS WITH Nr. of ABSTRACT :	GEN	PROF	SCI	SCO	GENORG	INST	FUNCT	SOURCES	HAND	CLASS	COD	EQUIP
the present state of biological literature (J10)							X					
magnetic drums,...punched ... tapes...,... patent documentation problems (J11)							X					X
logically cataloging knowledge... binary-octal designations (J12)										X	X	
Searching the chemical literature (J13)							X					
current research... in documentation (J14)			X									
Advances in EDP and information systems (J15)			X									
Technical communication in psychology (J16)							X					
evaluating scientific information retrieval systems (J17)			X									
(J18) : see (5) (?)									X			
Automatic categorization of chemical concepts and interrelationships (J19)										X		
Interrelationships as a basis for information retrieval (J20)										X		
preparing an inverted file (J21)												X

(6) (continued)

DESCRIPTORS : EXTRACTS WITH Nr. of ABSTRACT:	GEN	PROF	SCI	SCO	GENORG	INST	FUNCT	SOURCES	HAND	CLASS	COD	EQUIP
Patent Office searching (J22)									X (?)			
work at the... Laboratory of Machine Translation (J23)				X								
universal code of science (J24)										X		
Swiss documentation (J25)					X							
(J26) (?)			X									
(J27) (?)									X			
establishment... of a system for handling informal literature (J28)					X							
establishment of an international technical information system (J29)					X							
Directory of data processing service centers (J30)					X							
aeronautical documentation in the ...NATO countries (J31)					X							
(J32) : see (4)	X											
classification for the literature of documentation (J33)			X							X		
conference...for effective documentation systems (J34)			X									
Automatic selection and reproduction of industrial documents (J35)									X (?)			X

(6) (continued)

DESCRIPTORS : EXTRACTS WITH Nr. of ABSTRACT:	GEN	PROF	SCI	SCO	GENORG	INST	FUNCT	SOURCES	HAND	CLASS	COD	EQUIP
the Automatic Microfilm Information System (J36)									X (?)			X
design and construction ...of...a memory...a camera...an output system (J37)												X
Introduction to scientific documentation (J38)			X									
The Yugoslav Bibliographical institute (J39)						X						
Scientific literature use (J40)							X					
House organs and trade publications as information sources (J41)							X	X				
medical information in the Soviet Union (J42)					X							
Growth of chemical literature...Activity of Chemical Abstracts (J43)						X	X					
Soviet science information service...VINITI (J44)					X	X						
Hungary's network of technical libraries (J45)					X							

(6) (continued)

DESCRIPTORS : EXTRACTS WITH Nr. of ABSTRACT?	GEN	PROF	SCI	SCO	GENORG	INST	FUNCT	SOURCES	HAND	CLASS	COD	EQUIP
relationship of the library to society... information centers in Czechoslovakia (J46)	X				X							
documentation in West Germany (J47)					X							
Rapid Selector program of the Bureau of Ships (J48)									X (?)			
(J49) (?)			X						X			
microreproduction equipment... high capacity storage and retrieval and data management (J50)				X								X

2.2.2. Punching and editing.

A further step towards the setting up of the mechanized indexing system was the punching and editing of the texts. The 1121 abstracts of Janaske were punched on standard eighty-column punched-cards. The editing was carried out both before the punching and during the proof-reading of the cards. It was decided to punch the whole of the Janaske corpus in order to be able to prepare various types of word lists and indices.

Here I shall concentrate on the editing of the experimental corpus, and I shall not discuss the punching procedure or the production of indexes, since such a discussion is not necessary for the understanding of what follows. About the punching operations, however, I wish to mention three points, which are relevant to the design and organization of the system.

First of all, it was decided to limit the number of punching conventions to a minimum. For example, the German preposition "für" in "Institut für Dokumentationswesen" (J47) was punched "FUR" (see Appendix 3), and names which are written with capitals in Janaske such as "NATO" (J31) and "VINITI" (J44) were punched as if they had been written as "Nato" and "Viniti", that is, treated as capitalized words (see below). Similarly, the division of two abstracts of the experimental corpus into paragraphs (J34, J37) was ignored in the punching.

Second, all numerical indications (in a broad sense)
were punched with an asterisk preceding, as exemplified
below in (7). This was done with a view to a preliminary
simplification which is performed at the dictionary-lookup
stage : at this stage, all symbols preceded by an asterisk
are replaced by a single code, namely NUM. This undoubtedly
results in a loss of information but also in an important
simplification : it is no longer necessary to enumerate
all cardinal and ordinal numbers, and compounds containing
them, both in the form of digits (e.g. 5, 5th) and in the
form of the corresponding words (e.g. five, fifth). Such a
solution, of course, requires that the keypunch operator
knows the language : consider, for instance, the homograph
"second". The only numeral which was not punched with an
asterisk is "first" (not "1st"), because it can be used as
an adverb as in "First (of all)", "At first", etc.[(+)]"Binary-
octal" (J12) was also treated as an ordinary word. Examples
of the symbols treated as numerals in the experimental
corpus are given below :

<p align="center">(7)</p>

(a) <u>Examples</u> : *BILLIONS (J32),
    *EIGHTY-COLUMN (J35),*SECOND (J37),
    *SEVENTH (J15), *SIX-POINT (J5),
    *THREE-SYMBOL (J26), *TWENTY (J35),
    *0.3 (J37), *1ST (J38, J44),
    *104,484 (J43), *1957-1958 (J14),
    *2ND (J38, J50), *9800 (J43).
(b) <u>Dictionary-code</u> : NUM.

Third, all capitalized words were punched with an
asterisk following : for instance, AMERICAN* (J13, J28,
J33), WESTINGHOUSE* (J35). The asterisk is used in the
dictionary-lookup (2.3.) and proper-noun reduction (2.4.0.)
operations discussed below.

(+) Cp. the treatment of "ms": 2.3.1

The editing was also performed with a view to various preliminary simplifications. Not being an expert in programming, I am unable to say to what extent what was achieved by editing could have been done by means of appropriate computer programmes. What is clear, however, is that the problems which were solved in this study by editing are serious ones if one imagines their solution in a fully automatic system that would have to process language data as they appear in their printed form, with all its orthographic and typographic variations - both systematic and unsystematic : if, for instance, the texts were put in machine-readable form by means of some reading-machine.

Whatever the elegant solution to these problems may be, we found it expedient to solve them by editing, considering that we had only one year available to us to carry out our mechanized indexing experiment. Moreover, it was found that, even if the problems could be solved by stating additional rules in the system, such a solution would result in an unnecessary overburdening of the system.

Part of the editing was done by crossing out various types of information in the copy of Janeske used by the keypunch operators. The rest was done during the proof-reading of the cards, by punching again the cards containing data that we did not wish to be processed by the computer. The following types of information were changed or deleted :

(1) <u>Bibliographical information about the abstracted document</u>. Compare the facsimile reproduction of the first five abstracts in Appendix 4 with their edited version in Appendix 3. Thus, the name of the author (ACKOFF, R.L.) and <u>In....</u> in J2 were not punched. Nor was the indication © ADI appearing after some of the abstracts. Similarly, original titles in German, Russian, etc., were left out.

128.

(11) <u>Wrong or inconsistent spellings</u>. Consider, for instance,
"dessemination", second occurrence of "dissemination" in
J 5 , Appendix 4. An example of inconsistent spelling
is the capitalized "<u>The</u>" in the Janeske version of J30 :
"in Canada and <u>The</u> United States". This was normalized
into "the" and punched "THE". A similar normalization
was decided for the two occurrences of <u>(the)Federal</u>
<u>Government</u> in J5 : in Janeske, one is capitalized, and
the other is not (Appendix 4). The two occurrences have
been punched with asterisks in the edited version
(Appendix 3).

(iii) <u>Other normalizations and deletions</u>. Whereas there is
normally no space in print between various punctuation
marks such as comma, period, etc., or the mark of the
genitive ('s) and what precedes, these symbols were
punched as distinct symbols and could thus be treated
like words at the dictionary-lookup stage.
Various typographical signs which are listed below in
(8), were edited out :

(3)

(a) Quotation marks : "   "
(b) Points of suspension, as after etc.: etc...
(c) Various indications in enumerations :
    (1)(2)(3) and the equivalent Roman I. II. III.
(d) Dashes : - -
(e) Brackets.

Some dashes were punched as commas as in "the third
element - machine languages" (J24). "IBM 650" was
hyphenated and punched as a proper noun (J9). All that
was between brackets or dashes was either deleted or
changed, as is shown in (9) :

(9)

(a) computer research program of the Patent Office
    -- HAYSTAQ -- is described. (J18)
(a') COMPUTER RESEARCH PROGRAM OF THE PATENT* OFFICE*
    IS DESCRIBED .
(b) Two of the elements needed - terminological
    systems and symbolic languages - already exist.
    (J24)
(b') TWO OF THE ELEMENTS NEEDED ALREADY EXIST.

(c) in the pertinent geographical region (German,
French, Italian). (J25)
(c') IN THE PERTINENT GEOGRAPHICAL REGION : GERMAN*,
FRENCH* , ITALIAN*

Lastly, various bibliographical data appearing in the
texts of some abstracts were edited out : for instance,
", by B.E. Holm, p. 64-74" at the end of J15. In only
two abstracts (J15 and J33) out of fifty, the various
normalizations and excisions reported in this section
have resulted in drastic modifications of the printed
versions of the texts. The titles and texts of these
two abstracts are quoted below in (10) as they appear
in Janaske, and this unedited version can be compared
with the edited one given in Appendix 3 (incidentally,
note a punching error in "AMA's", J15). Note that most
of these drastic changes are due to the occurrence of
bibliographical information in the printed version of
these texts. In fact, in the absence of such information,
little or no editing was found necessary.

(10)

(a) Advances in EDP and information systems.

Papers in this volume are "based on material orig-
inally presented at AMA's Seventh Annual Data
Processing Conference, held in March 1961".
Partial contents : Advances in information
retrieval and data acquisition: I. Progress in
the design of information retrieval system, by
M. Taube, p. 51-63; II. Improved information
storage and retrieval systems, by B.E. Holm,
p. 64-74.

(b) A suggested classification for the literature
of documentation.

The classification scheme presented here had its
origin in a class project undertaken by students
at the University of Chicago, Graduate Library
School during the spring quarter 1960. A
description of the principal features of the
classification with reference to David J. Haykin's
six basic building rules for special classifica-
tion is followed by 2 appendices: I. Classified
Guide to American Documentation 19.0-1960, vol.
I-XI; II. Suggested classification for the
literature of documentation.

2.2.3. System design.

In this section, I wish to discuss two points which concern (i) the theoretical foundations and (ii) the general strategy of mechanized indexing. My reason for bringing the points together here is that are related as far as the historical background of my experiment is concerned, and they both concern decisions which were crucial for the design of the system as it was actually set up.

Concerning the theoretical foundations, I must stress the fact that during the actual design and development of the system a number of conditions were absent which, had they been met, could have changed the theoretical orientation of my experiment; in fact, some of these conditions are still not met at the time I am writing this thesis.

First, it was not possible for me to adopt and explore the methods proposed by the Gardin group and reported in Bély et al. (1970), even though a pre-publication version of their report was available to me in due time. The research of the Gardin group hinges on semantic representations in terms of the SYNTOL model (Cros et al. 1964 and 1968), and even the Gardin group which has been much concerned with the field of documentation has so far not attempted to develop SYNTOL descriptors or extensions of the SYNTOL syntax for the field with which I am concerned (see N. Gardin and F. Lévy 1967). I myself have not attempted to analyze the Gardin-Lévy concordance in terms of the SYNTOL model (my only attempts towards an analysis of this concordance appear in Part III).

Second, most of the other linguistically oriented
studies in automatic indexing which I have heard of were
not available to me in detailed reports during my experiment.
Even if they had been, they could not have been exploited
in my work. Some of them, in particular the research of the
Salton group with the SMART system and Sager's String
Program, would have required computer programmes which have
never been available to me : for instance, the Kuno-Oettinger
analyzer used in the SMART system (Salton in Hays 1966, 180).
Other American projects had published studies which were
either based on a very crude syntax (Stone at al. 1966), or
were - and, as far as I can see, still are - based on meta-
theoretical assumptions which are not compatible with those
presented in Part I - in particular, with my concern with
a fairly large corpus of actual texts in relation to an
actual list of individ specific descriptors. The following
is quoted from the conclusions of a state-of-the-art report
by Simmons entitled "Natural Language Question - Answering
Systems : 1969"; the two publications (out of a list of 79)
mentioned by Simmons as going beyond sentence boundaries
have never been available to me :

> Significant weaknesses are still prominent. All
> existing systems are experimental in nature, small,
> and corebound. None uses more than a few hundred
> words of dictionary or a small grammar and semantic
> system. None can deal with more than a small subset of
> English strings. Deductive operations, though undeniably
> powerful, still generally lack adequate heuristic
> controls to limit the extent of searching an infinite
> inference tree. Little has been done so far to
> incorporate inductive inference procedures. Few systems
> (see Charniak and Wilks) go beyond sentence boundaries
> in their analyses, and generally acceptable methods
> for anaphoric analysis and the discovery of prohominal
> reference have not yet been developed. Such subtleties
> as the relativity of adjectives and adverbs, thematic
> sequence, metaphor, etc., have still to be explored.
> (Simmons 1970, 28-29).

Third, Part III of this thesis was prepared __after__
my mechanized indexing experiment was completed, and most
of the theoretical sources on which that part of my thesis
is based appeared at the time the system was developed, or
were not available to me in due time for them to affect
the design of the system.

As to what I have called the general strategy of
mechanized indexing, the decisions which underlie my experi-
ment are best explained with reference to those that underlie
the experiments of the Gardin group reported in Bély __et al__.
1970, since theirs is the only automatic indexing system I
know that is based on metatheoretical assumptions similar
to mine. The main issue here, I believe, is the respective
place of semantic and grammatical considerations in the
actual development of a computerized indexing system. Since
any actual system necessarily emphasizes one aspect at the
expense of the other, I feel justified in reducing the issue
to a conflict between two rival strategies. One is best
explained by quoting the first sentence of a paper by Sagor;
the other, by quoting from a paper by J.C. Gardin, which was
read in 1964, at a time when the Gardin group based their
automatic indexing system entirely on semantic information.

Writers of programs for handling natural language
material, whether for mechanical translation, automatic
indexing, or other purposes, have learned (sometimes
the hard way) that without grammatical analysis their
efforts at processing language data reach an early limit.
(Sagor 1967, 153).
... observons ici encore l'analogie avec la traduction
mécanique, où l'on distingue de la même manière entre
la mise en oeuvr. d'un dictionnaire automatique pour le
choix des mots, et l'application d'un algorithme
d'analyse et de synthèse grammaticalespour l'expression
des relations logiques entre ces mots.
C'est sans doute en raison de cette analogie que
l'on a souvent abordé "l'analyse automatique", en
documentation, par le biais de l'analyse syntaxique des
textes naturels, le raisonnement implicite paraissant
être que celle-ci était en tout état de cause nécessaire
à celle-là. Rien cependant n'est moins évident : en
premier lieu, il n'est pas rare que l'on se cont nte
en pratique d'une représentation par descripteurs seuls,

sans indications syntaxiques, auquel cas l'algorithme
d'analyse grammaticale perd une partie de son objet
(...) En second lieu, lorsque des indications syntaxi-
ques sont requises, elles sont souvent en fait si
pauvres (par exemple, un simple repérage de liens non
interprétés entre certains descripteurs pris 2 à
2 ou n à n) que l'on peut se demander si l'analyse
logique intégrale des textes d'entrée n'est pas un
outil disproportionné à la tâche. Enfin, et surtout,
il est connu que la reconnaissance des structures syn-
taxiques ne permet elle-même de résoudre qu'une partie
des difficultés de traduction, au point que l'usage
s'est établi d'annoncer toujours une phase ultérieure
d'<u>analyse sémantique</u>, qui devra "compléter" la précé-
dente, "supprimer les ambiguïtés résiduelles", etc.
Malheureusement, cette phase finale n'est jamais décrite,
autrement que dans son principe; et si ce dernier paraît
en général fort simple - il s'agit toujours, sous une
forme ou une autre, de la "résolution par le contexte",
déjà citée - l'application n'en demeure pas moins pro-
blématique (...) les données linguistiques et extra-
linguistiques nécessaires pour mener à bien cette analyse
sémantique risquent en fait d'être si complexes que l'on
est dès lors tenté de les employer non plus comme
"complément" des données grammaticales, après une
analyse syntaxique déjà coûteuse, mais comme "substitut"
de celles-ci, pour mener à bien l'interprétation
syntaxique généralement sommaire que l'on vise en
documentation. (J.C. Gardin 1966, 120).

The Gardin team have since then made use of more and
more grammatical information. Their system, as described
in the final report by Bély et al. 1970, is based on a
complex interaction of grammatical and semantic analysis
procedures, at the stages of both "lexical" and "syntactic"
indexing. Thus, at the 'lexical" stage, some types of word
groups are translated into SYNTOL descriptors partly on the
basis of grammatical procedures : one serves to select the
appropriate part-of-speech category (e.g. adjective or noun
for "complexe" : op. cit., 27), and another, which works
in combination with semantic analysis, establishes some
syntactic relations : for instance, the relation of an
adjective to the noun it qualifies (op. cit., 31).

At the "syntactic" stage, which serves to establish the
nature and orientation of some SYNTOL relations holding
between the descriptors obtained at the "lexical" stage,
both parts-of-speech categories and some syntectic relations
are used once again together with semantic information.
Thus, the machine can operate on the basis of syntactic
templates ("schémas syntaxiques") for establishing a given
SYNTOL relation from the various possible grammatical
structures that correspond to it : for instance "... provoque
...", "... qui provoque...", "... que provoque...", "...
provoqué par...", etc. (op. cit., 103). This grammatical
information can be complemented by semantic information,
obtained by looking up tables (a so-called semantic or
notional network) specifying a priori the SYNTOL relation
that can hold between two descriptors and between the
corresponding natural language words : for instance, in an
expression like "the effect on A of B", the relation
established between A and B will not be the same for A =
sleep and B = an animal as for A = an animal and B = some
chemical, and no grammar is able to predict such a fact.

When designing my system, I made the following decisions
which can be regarded as a compromise between the extreme
'grammatical' approach represented by Sager's statement, and
the exclusive reliance on semantic analysis as in the
first experiments of the Gardin group :
(1) I decided to give primacy to surface grammar and found
    that my system should be based essentially on rules
    that make use of parts-of-speech categories of the type
    proposed by Harris in string theory, and similar to
    the grammatical codes used by Bély et al. 1970 and many
    others (in particular Hornby's "Patterns" : Hornby 1954).
    I also found that Harris's string theory was the most
    explicit treatment available of surface facts of concatena-
    tion, and the one which, because of its explicitly
    algebraic formulation, best lends itself to computational
    analysis, and to an 'autonomous' treatment of what Chomsky
    regards as (derived) surface structure.

(11) I decided to define the tasks of the various 'grammars'
in the system in terms of grammatical relations (including
adjunction), but <u>not</u> to let the machine establish, or
operate with, symbols for these relations, such as the
<u>Subject+Verb+Object</u> formula which defines the notion of
center string (Harris 1965, 24; cp. Mély <u>et al.</u> 1970). My
first reason for leaving grammatical relations implicit
is a practical one : I felt that this additional dimension
could not be introduced into the system within the time
available for the experiment. Besides, as pointed out in
Part II, the concept of center string is of questionable
semantic validity. Moreover, the concept breaks down in
"semi-articulate" ("Thank you !") and "inarticulate"
("Thanks !") sentences (Jespersen 1965, 308; see 3.2.3.).

(iii) Subcategorization (2.3.) is handled by dictionary codes
which can be changed or improved and is thus left open
to further investigation. Similarly, the concatenations and
rewritings in the various 'grammars' of the system are
based on manipulations of codes which could be improved
(or even changed altogether, if texts in another field were
to be processed) WITHOUT affecting the general organiza-
tion of the system or requiring changes in the computer
programmes. This strategy by which (for lack of a better
solution) some crucial theoretical questions are left
open is, I believe, a sound one in computational work,
because one cannot afford to take the risk of wasting
programming and machine time by setting up a system
hinging on aspects of linguistic analysis that obviously
require further research (see Part III).

This, I must stress, should not be interpreted as a criticism of Bély et al.'s experiment. Their situation was different from mine in that they had both SYNTOL descriptors and the SYNTOL syntax around which they could organize their experiment. They have eveloped a most interesting approach to problems of discourse structure (in particular, their use of various types of contexts to establish some descriptors, and their "raccordement" : Bély et al. 1970, 30 and 117). But their approach is based on the SYNTOL framework which has so far not been applied to the field of documentation.

(iv) My system has been defined as one which works from (object-language) surface structure to (metalanguage) surface structure, and, more precisely, as one which reduces the longer strings of the object-language to the shorter strings of the metalanguage. The essential function of the system is therefore to reduce long strings to short ones. This means that the longer and often more explicit expressions are (progressively) replaced by equivalent short and mostly less explicit, symbols. Consequently, by analogy with TG theory, the system can be described as one that transforms texts into semantically relatable noun-phrases. The reduction process is parallel to various similar processes which can or could be stated about the object-language : for instance, a long noun-phrase with a relative clause such as "the... Center... which was established for the purpose of collecting, processing, storing, and retrieving scientific information"   is relatable and reducible to the noun-phrase "the documentation center" which, being shorter, is closer than the other to the meta-language descriptor. The purpose of the system presented

below is to mechanize such reductions on the basis
of concatenation rules, and the purpose of Part III of
this thesis is to throw some light on the semantics that
underlies these reduction processes, and the converse
possibility of expressing in a sentence or in a text what
can partly be expressed by a shorter string.

2.3. SUBCATEGORIZATION AND DICTIONARY-LOOKUP.

2.3.1. Introduction.

The dictionary presented in this chapter was set up in
order to enable the computer to replace over one thousand dif-
ferent word tokens occurring in the experimental corpus by
codes consisting of no more than 15 characters. Not all of
these thousand words were assigned a specific code in the
dictionary. Thus, most of the capitalized words, which were
punched with an asterisk following, were left uncoded and
were replaced mechanically by the code NOT CODED (see the
treatment of proper nouns below, in section 2.4.0.). Similarly,
the words that were punched with an asterisk preceding, that
is, numerals and words relatable to them, were automatically
replaced by the code NUM. In this chapter, I shall not be
concerned with the words replaced by NOT CODED and by NUM.
I shall only deal with the others.

As pointed out in 1.3.2., the dictionary was set up on
the basis of proposals made by Salkof and Sager 1967 concerning
a restrictionless string grammar. This is a string grammar
whose rules state dependencies of subclasses or subcategories.
These dependencies are intended to be restrictionless, as
against the dependencies of more general classes like those
used by Harris in the first version of string theory (Harris
1965): for instance, A for adjective and N for noun. Thus,
one can assume that a string formula AN is not restrictionless
in the sense that any adjective cannot co-occur with any noun.
A restrictionless string grammar, then, can be defined ideally
as one whose rules or formulas have no exceptions.

The purpose of this chapter is to explain how I set up the subclasses or subcategories for my experiment, what kind of subclasses were set up, and what are the consequences of my approach.

The task of setting up a dictionary is by no means an easy one. As far as the experiment under discussion is concerned, there were some additional difficulties. The number of words to be examined was large (over a thousand) and the time available for doing this was short : since dictionary-lookup comes first, it was the first operation that had to be programmed and tested, so that only a few months were available for the setting up of the dictionary, and even some of this time had to be devoted to the general design of the system and planning of the experiment. The dictionary codes which were eventually set up are therefore extremely tentative, as will appear from the ensuing account. Moreover, I have made generous use of a convention (the period symbol, explained below) which amounts to saying that for a given word I have been unable to set up any kind of motivated subcategory. This, I suggest, was the most reasonable thing to do considering the time that could be devoted to the dictionary codes.

In addition, I must point out that I am unable to explain how I arrived at my dictionary codes in terms of any formal procedure, or to present arguments for or against my codes. Such a procedure does not exist, as far as I can see : Katz and Postal, for instance, would equally be unable to explain how they establish relations between markers on the basis of the category inclusion relation mentioned in the passage quoted below. In my opinion, the method that would come closest to a formal procedure for setting up lexical subcategories would have to make use of Lyons's and Leech's work on implication

(see below, section 3.3.4.) or of Harris's proposals about
the concept of "vocabulary classifiers". "The vocabulary
classifiers are the last N in sentences of the form X is a
(member of, or : case of, etc.) N (...) E.g. (...) a salmon is
a fish" (Harris 1968, 166-7). What is by no means clear in
Harris's suggestion is the nature of the predicate or relation,
in the logician's sense of the word : note the "etc.".
Moreover, it seems to me that Lyons's implication relation is
more general and interesting than Harris's N is Nol proposal,
because implication holds between sentences and can therefore
involve other grammatical classes than N.

What I have NOT attempted to do in my experimental
subcategorizations is best explained by referring to Katz
and Postal 1964. I have not attempted to set up dictionary
entries which are claimed to be valid in the whole of a
language or to reflect "linguistic universals":

> The formulation of a dictionary for the semantic
> component of a particular language can be greatly
> economized by taking advantage of a relation between
> certain pairs of semantic markers. The relation
> that serves this purpose is the category inclusion
> relation which holds between a pair of semantic
> markers when the category represented by one is a
> subcategory of that represented by the other. For
> example, the semantic marker (Human) represents
> a conceptual category that is included in the
> categories represented by (Animate), (Higher Animal),
> (Physical Object), etc., but the category that the
> semantic marker (Physical Object) represents is not
> included in any of these aforementioned categories.
> These category inclusion relations will be specified
> within the general theory of linguistic descriptions
> as part of that theory's statement of the semantic
> concepts that are linguistic universals when such
> a statement is a true generalization about the
> structure of the dictionary for every linguistic
> description. (Katz and Postal 1964, 16).

Following suggestions made by Harris 1968, and contrary
to Katz and Postal's emphasis on universal or language-
independent properties of the dictionary, I have attempted
to set up subcategories that capture what is specific or
distinctive about the English of abstracts in the field of
documentation, _and_ about categories of abstracts as esta-
blished by indexing.

In short, then, the application made in this study of
the concept of restrictionless string grammar has two aspects.
One is the aspect envisaged by Harris, Salkof and Sager and
concerns the statement of dependencies between subclasses
of words by means of string formulas. This aspect is the one
which is most directly relevant to computer recognition : once
the dictionary-lookup programmes have replaced the text words
by codes standing for subcategories, the computer can scan
the strings of codes in order to recognize those strings
which match exactly a rule r string formula stored in its
memory, and the recognition can proceed without any restric-
tions.

The other aspect concerns the way in which the descriptors
(and some additions made to them, see below), and therefore
also the concatenation rules which serve to eventually reduce
the texts to descriptors, determine the nature of the
subclasses or subcategories which were adopted for the codes.
To take an example based on the above passage from Katz and
Postal, the word "psychologists" (J16) in the corpus under
investigation need not and should not be characterized by
markers such as (Animate), (Higher Animal), (Physical Object).
The marker (Physical Object), for instance, would presumably
be 'relevant' to an analysis of say, "psychologists who weigh
more than 200 pounds" which can be assumed to be impossible

in abstracts on documentation, and such a marker is irrelevant
to the analysis of a sentence (and of the abstract) in which
the word actually occurs in the experimental corpus :
"psychologists should attend immediately to the problem of
encyclopedic organization of their knowledge..." (J16).
I do not mean to suggest that the theoretical status of what
I have just called 'relevance' is clear. All I wish to do
here is to raise my problem on a sound empirical basis,
rather than on an obviously incorrect one. To me, at least,
such concepts as (Physical Object), (Higher Animal), etc.,
cannot be associated with "psychologists" in the above
sentence and cannot be automatically assigned by the theory
to the semantic representation of this sentence, and it is
important to go out from such facts.

Details about my attempt to let the dictionary codes
be determined by the indexing will be given in the next
section (2.3.2.). In particular, I shall explain my attempt
to let the degree of generality of the subcategories be
determined by the indexing, rather than have for a given
word an indeterminate number of categories included in one
another, as Katz and Postal would propose : I assume that,
for instance, a machine or theory which computes "THE THING
IS ON THE THING" from the sentence "The cat is on the mat" is
too indiscriminate for the distinctive purposes of indexing.

With a few exceptions to be mentioned below, the codes
that were set up for the words occurring in the experimental
corpus consist of two parts. These two parts correspond to
the kind of notation proposed by Harris and mentioned in
1.3.2. for sentences such as "The polypeptides were washed
in hydrochloric acid". The first part represents categories
"recognized in the whole language", such as the following in

Harris's formula for the above sentence : <u>N is Ven in N.</u>
The second part corresponds to the subscripts (for instance,
$\underline{N}_{\underline{mol}}$ and $\underline{N}_{\underline{sol}}$) which Harris uses to "name particular sub-
classes" "recognized (...) only in the science sublanguage"
(Harris 1968, 153). Like Harris's notations, my dictionary
codes are abbreviations of ordinary words. Such abbreviations
were necessary to stick to the 15-character coding format
just mentioned.

To simplify the presentation, I shall henceforth refer
to the first part of my codes as a <u>category</u> (cp. Harris's
<u>Ven</u>) and to the second part as a <u>subcategory</u> (Harris's sub-
scripts). To make the codes used in the experiment more
readable, these two parts were separated by a bar (/). Thus
ACTIVITIES was coded NPL/ACT.; NPL stands for Noun Plural,
and the period (.) indicates that the notation is an abbre-
viation of the actual graphical form of the coded word.

The use of the period symbol in the experiment was not
always justified or consistent. I believe that it was,
however, in two types of situations. One can be exemplified
by the name "Chemical Abstracts" (J43). It is essential for
the system to identify and analyze this name correctly in
order to be able to index abstract J43 by means of the
descriptor STI INSTITUTIONS. The period symbol assigned to
the codes for CHEMICALx and for ABSTRACTSx serves to express
the fact that these codes stand for the words "Chemical" and
"Abstracts" and nothing else.

The other important use of the period symbol is the
following. Whenever I felt unable to assign a subcategory to
a word (and this turned out to be very frequently the case)
the graphical shape of the word was used (occasionally

in abbreviation) as the code and marked by a period :
for instance, the code NSG/COR/E. for the word "core".In
the following section, I shall not give a complete list of
the words coded in this fashion. Such a list would be of no
interest, since I am unable to say why I could not subcate-
gorize the words in question. I shall just mention a few other
examples here, because they will appear in an illustration
given below in 2.5.1. : "service(s)" was just assigned the
subcategory coded as SER/VIC.; "operation(s)", the subcategory
OPE/RTN.; "science", the subcategory SCI.; and "information",
the subcategory INF..

2.3.2. Nouns, adjectives, and verbs.

The categories assigned to nouns are given below in (11), column 1; in column 2, I indicate and exemplify what the code stands for. Note that these types of codes are in agre ment with the concept of restrictionless grammar and that, thanks to such codes, it is possible to do without a procedure for morphological analysis. All words which can only be adjectives were labelled ADJ .

(11)

NAA	proper noun or adjective : RUSSIANx (e.g. in Russian)
NAM	'name',i.e. proper noun;or adjective:AFRICAx,SWISSx
NPL	noun,plural: ACTIVITIES
NSG	noun,singular: EFFORT
NSP	noun, singular or plural: DATA (only example)
NXX	noun, or some other category: SUBJECT
N2G	noun,or ing-form: BEGINNING
N2P	noun plural, or verb in 3rd person: WORKS
N2S	noun singular, or verb in another person of the present tense : WORK

The grammatical status of homographs such as N2P was established by rewriting rules : for instance, the concatenation of "this" with the following "reviews" serves to establish that "reviews" is a verb in J2 (see G.2., below).

Homographs which were coded under the general heading Adjective, are of the following types :

(12)

A2N	adjective or noun: FUTURE, INDIVIDUAL, etc.
A2V	adjective or verb: DIRECT, COMPLETE, PRESENT (which could have been coded A3),etc.

Examples of the effect of indexing on the <u>subcategories</u>
chosen for the experiment are the codes GEO/NAT, standing
for 'geographical, national', and GEO/INT, standing for
'geographical, international'. GEO/INT was assigned to
words such as the following : "<u>foreign</u>" (ADJ/GEO/INT),
"international" (ADJ/GEO/INT.), "<u>Europe</u>", "<u>Africa</u>", "<u>Nato</u>"
(NAM/GEO/INT), "<u>abroad</u>" (N+P/GEO/INT:N+P stands for prepo-
sition+Noun), etc. GEO/NAT was assigned to such words as
"<u>national</u>" (ADJ/GEO/NAT.), "<u>country</u>" (NSG/GEO/NAT), "<u>Swiss</u>",
"<u>Switzerland</u>" (NAM/GEO/NAT), etc. Now, as appears from
2.2.1., heading 5 of Gardin-Lévy ("<u>General organization of</u>
<u>STI</u>") is the best represented descriptor in the experiment :
14 abstracts out of 50 are represented by means of this
descriptor, and part of its definition, which is quoted in
Appendix 1, reads : "Ensemble des organismes qui forment
l'appareil général de l'IST, <u>dans un ou plusieurs pays</u>,
un ou plusieurs domaines d'application".The 'geographical'
codes given above and their degree of generality were
adopted to match the underlined portion of this definition.
Several codes set up for the final reduction of the abstracts
to a single code (in G.8) consist of two parts : one
symbolizing Gardin-Lévy's heading and the other some addi-
tional information. In the case of the abstracts described
by "<u>general organization of STI</u>", the final code consisted
of the indication NETWRK (standing for the above heading:
'network of STI institutions') and, if necessary, of the
indications NATION ('national') or INT ('international')
next to other additional indications. Thus, J25 which is
subtitled "Guide to <u>Swiss</u> documentation" was eventually
reduced to the code NETWRKNATION, and J29 which concerns
"a proposal for the establishment of an <u>international</u>
technical information system" was reduced to NETWRKFUTINT
(where FUT stands for 'future').

Also in connection with the descriptor NETWRK, I
wish to mention the subcategory INS standing for 'institu-
tion'. When a noun thus subcategorized is plural, it
directly matches the plural in Gardin-Lévy's definition of
"Network of STI institutions". Consider the following
examples : "organizations" (J14, J42), "bureaus" (J30)
(both coded NPL/INS), "institutions" (J2, J25, J38, J47 :
NPL/INS.), "libraries" (J25, J38, J40, J45, J46), "archives"
(J25) (both coded NPL/INS/DOC, i.e. 'documentation institu-
tions'), "branches" (J46; NPL/INS/PAT, i.e. 'part').

Another example of the way in which the indexing
determined the nature of the dictionary codes is the abstracts
described by the heading "Functioning of STI networks",
which is also a frequently used descriptor in the experi-
mental corpus (Gardin-Lévy's heading N° 7 : eight abstracts).
In the final rewriting for the abstracts, this heading was
symbolized as FLO, standing for 'flow of information'.
Part of the definition for this heading reads : "Circula-
tion de l'IST (...) appréciations sur le fonctionnement
des réseaux d'information considérés au groupe 5". Now,
the reference to the field of application in the definition
of Group 5 just quoted ("un ou plusieurs domaines d'applica-
tion") turned out to be one of the characteristics of
abstracts described by FLO. Examples are the titles of
J4 and J13 : "Searching medicinal chemical literature" (J4),
"Searching the chemical literature" (J13). When applicable,
this reference to the field was added to FLO in the final
rewriting of the abstracts in the form of the code DIS,
standing for "in a particular discipline". This was done
for J4, J10 ("Storage and retrieval of biological informa-
tion"), J13, J16 ("Technical communication in psychology")
and for J43 ("Growth of chemical literature"). This symbol
DIS served to subcategorize adjectives such as "chemical",
"medical", etc. (ADJ/DIS), "industrial" (ADJ/DIS/BUS, i.e.
'business'), nouns such as "chemicals" (NPL/DIS/TOP,

i.o. 'topic of a disciplino' as in "a manufacturer of fine chemicals"; J4), "specimens" (coded like "chemicals"; "storage and retrieval of biological specimens" in J10), "compounds" (coded like "chemicals"; J22), "psychologists" (NPL/DIS=AGT, i.e. 'agent'; J16), "psychology", "chemistry" (MSG/DIS), etc.

The topic FLO can also be expressed by references to the documents whose 'flow' or 'circulation' is considered, as in "patent documentation" (J11), "scientific literature use" (J40), and in "acquisition and handling of house organs and other trade publications" (J41). The code ITM ('item'), with occasionally an additional code such as TYP ('type') and PAT ('part'), was used to subcategorize words for various sorts of documents; the code ITM was used in order to avoid the code DOC which is used for other purposes explained below. Examples of words subcategorized by the code ITM are : 'paper' and 'article' (ITM); 'item' (ITM.); 'pages', 'appendix', 'chapter' (ITM/PAT); 'disclosures', 'patents', 'memoranda', 'drawing', 'organ', 'journal', 'prescription' (ITM/TYP); 'collections', 'holdings' (ITM/VAG, i.e. 'vague', standing for 'mass noun') and 'literature' (ITM/VAGLIT.).

One last way in which indexing determines the nature of the dictionary codes is worth mentioning here in connection with certain nouns and adjectives. It has to do with various losses of information which take place in indexing. Thus, certain words are, or seem to be, irrelevant to indexing. However poorly understood this loss of information may be, and however dangerous it may be to effect this loss of information in the dictionary codes, it was decided to do so for a number of words. Thus, adjectives like "net", "fine", "broad", "basic", "modern", were

assigned the subcategory QAL standing for 'some qualitative
indication', and adjectives such as "finite", "entire",
"many", "linear", "countless", "low-cost", the code QAT
standing for 'some quantitative indication'. Occasionally,
a code was added to these general subcategories, but these
additional codes are in general too arbitrary to be worth
mentioning here. Similarly QAL was used to code nouns like
"advantages" and "aspects", and QAT for nouns like "prices",
"rates", "fact". No QAL or QAT codes appear in the final
rewritings for the abstracts established in G.8 and explained
in the relevant section of the next chapter. Nouns like
"behavior", "efforts", "tasks" were indiscriminately sub-
categorized as ACT, a code also use for "activities" (ACT.),
and nouns like "consideration", "preparation", "creation",
"generation", as DEV, the subcategory used for "development"
(DEV.). Like QAL and QAT, the codes ACT and DEV do not
appear in the final rewritings of the abstracts.

A particularly important subcategory, which was used
for quite a number of verbs and cognate nouns, is the code
DOC. It served to subcategorize words for various activities
in the field of documentation : "acquisition" (DOC/COL),
"documentation" (DOC.), "classification" (DOC/CLA.),
"dissemination" (DOC/DIS.), "retrieval" (DOC/RTR.), "storage"
(DOC/STO.), code (N2S/DOC/COD.), "search" (N2S/DOC/SCH.),
"file" (N2S/DOC/STO), "abstracting" (V.G/DOC/ABS.),
"cataloging" (VNG/DOC/CAT.), "collecting" (VNG/DOC/COL.),
"gather" (DOC/COL). As will be noted, the above codes make
generous use of the period symbol. This means that I did
not wish to commit myself to any systematic subcategoriza-
tion of the words in question during the mechanized indexing
experiment. I felt that the problems involved could be more
usefully examined in the semantic part of my investigation
(Part III).

Finally, I wish to mention the subcategory STU, which
was used to code various words which serve to introduce the
topic. of the abstracts but of which I assumed that they
do not serve to express these topics. These words, which
are also discussed in Part III of this study (chapter 3.2)
include the following : "deliberations" (STU), "observa-
tions" (STU/OBS.), "discussion" (STU), "investigation"
(STU), "description" (STU/DES. in which the specification
DES. is intended to predict the special use of "describe"
as in the word 'descriptor'), "evaluation" (STU/QAL),
"studies" (N2P/STU.), "measures" (N2P/STU/QAT), "traced" and
"outlined" (VED/STU). Another subcategory which was assumed
to be related to STU is SAY, which was assigned to such verbs
as "purport", "conclude", and "state".

I shall not present any further subcategory set up
for the experiment, either because I feel that they are too
tentative and arbitrary to be of interest, or because
they are not dealt with in Part III. All the subcategories
that are worth mentioning for verbs have been mentioned
above in connection with cognate nouns.

The grammatical categories used for verbs are listed
below:

(13)

VED	For instance, UNDERTAKEN, ACHIEVED
VNG	For instance, DEVELOPING, PREPARING
VPL	For instance, REQUIRE, MAKE
VSG	For instance, FEELS, SHOWS
VXX	V, VED, VPL or Noun : CUT (only example)

The codes for forms of 'to be' and a few other verbs
are presented in the next section.

2.3.3. Other categories and codes.

Special codes were set up for categories other than
Nouns, Adjectives and Verbs, and for words which have special
properties. These special codes are listed below in (14).
The codes consisting of X plus the graphical shape of the
coded word (see below) were set up for the same reasons as
those which dictated the use of the period symbol : either
they occur in phrases or phrase-like constructions (cp.
"as well as" and "Chemical Abstracts"), or no category
can be safely assigned to them ("as" is a good example),
or both ("as well as").

<div align="center">(14)</div>

A	For A and AN
THE	For THE
ALL	For ALL
ANY	For ANY
ADV	For adverbs such as ALREADY, ALSO
ADX	For adverbs such as ONLY, PARTICULARLY
CC	For semi-colon
CR	For OR
C1	For comma
C2	For AND
KL	For colon
S	For period
P...	Prepositions. For instance PBY, for the preposition BY.
PS	For 'S, the mark of the genitive
X...	See below.

A difference was made among adverbs between ADV,
ADX, and adverbs that are subcategorized because of their
assumed importance for indexing. The latter include :
"manually", "automatically", "mechanically", and "logically".
Some adverbs were coded ADX because observations showed
that they could be followed, or preceded (or both) by punc-
tuation marks such as comma. These adverbs were deleted by
concatenation rules in G.0 (see below). They include :
"particularly", "however", "especially", and "today". The
adverbs coded ADV were deleted by a special subroutine

which will be discussed later. The purpose of the codes
CC, CR, C1, C2, KL, and S is also explained below in
Chapter 2.4.

The words whose code consists . of X followed by the
word in question are listed below (15), without the X and
in alphabetical order. They often belong to more than one
category. The categories they belong to include subordinate
conjunction, adverb, verb (<u>to be</u>, <u>to have</u>, modals in their
various forms), possessive and demonstrative pronoun and
adjective, personal and relative pronoun, and negation.

(15)

ABOUT	NOT
AFTER	OFF
AHEAD	ONE
ALTHOUGH	ONES
ARE	OTHER
AS	OUR
BE	OUT
BECAUSE	OVER
BEEN	SHOULD
BEFORE	SINCE
BEING	SO
CAN	SOME
EACH	SUCH
ETC	THAN
FAR	THAT
HAD	THEIR
HAS	THEM
HAVE	THEN
HAVING	THERE
HE	THESE
HIS	THEY
HOW	THIS
IF	THOSE
IS	UNLESS
IT	US
ITS	VERY
LEAST	WAS
LESS	WELL
MAY	WERE
MERELY	WHAT
MIGHT	WHATEVER
MORE	WHEN
MOST	WHICH
MUCH	WHO
MUST	WILL
NO	WOULD
NOR	

2.4. DESCRIPTION OF THE SYSTEM.

2.4.0. The processing of proper nouns.

By proper noun is meant here not only nouns but, by
extension, any capitalized word. The procedure set up for
the processing of proper nouns is based on the following
considerations :
(i) there exists to my knowledge no grammar of proper nouns.
(ii) the list of proper nouns is open-ended.
(iii) a distinction needs to be made among proper nouns
between those that are required for the purpose of
indexing and those that are not, however difficult such
a distinction may be.

On the basis of these considerations, I set up a
procedure for the treatment of proper nouns which I think
is the most general and most fully automatic procedure in
the system. It works as follows. Capitalized words, which
have been punched with an asterisk following, can be given
two types of codes in the dictionary-lookup operation.
Some are assigned a code standing for a specific semantic
class : for instance, "Japan" is coded NAI./GEO/NAT standing
for Proper noun, geographical, national. Others are replaced
by NOT CODED, which means that the dictionary does not
contain a code for them : for instance, "Context" is replaced
by NOT CODED.

When a capitalized word occurring with no other capital-
ized word has been replaced by NOT CODED, THE+NOT CODED
is in turn replaced in the proper noun reduction procedure
by the code NPP standing for proper noun : for instance,
in "the BSA" (J 16) BSA× is replaced by NOT CODED, and
THE NOT CODED   by NPP. Otherwise, if the capitalized word
has been replaced by a semantic code, this code is not
processed by the proper noun procedure : for instance,
"U.S." occurring in "particularly among U.S. scientists and
engineers" (J 5).

Compound proper nouns, that is, those consisting of more
than one capitalized word, are handled as follows. If at
least one NOT CODED appears in the corresponding string of
codes obtained by dictionary-lookup, all this string,
including the definite article that precedes, is replaced
by NPP : for instance, "the National Science Foundation"
(J 2). If all the constituents of a compound proper noun
have been replaced by semantic codes, these codes are handled
by the rest of the system; this is the way I have treated
"American Documentation" (J 33), "Chemical Abstracts" (J 43),
"the Federal Government" (J 5) and various names of countries
such as "the German Democratic Republic" (J 7), because
I assumed that they were important for indexing.

A further distinction needed to be made among compound
proper nouns between those that consist of only capitalized
words following the article ("the National Science Foundation")
and those in which some non-capitalized word or symbol occurs
between the capitalized words : for instance, "the Japan
Information Center of Science and Technology" (J 1).
In order to also reduce the latter type to NPP, the programme
was written so that the machine would disregard the occurrence
of the non-capitalized 'words' by looking up a list of these
words. The list for the experimental corpus is given below
in (16). A summary of the proper noun procedure is tabulated
in (17).

(16)

```
,
;
)
and
at
for
fur (J47)
in
of
the
```

(17)

Type :	string with at least one NOT CODED		Yes : replaced by	No : replaced by
One-word			NPP ("the BSA", J16)	semantic code ("IBM-650", J9)
Many-word	with a 'word' of list (16)		NPP ("the Board of Scientific Affairs, J16)	semantic codes
	without		NPP ("the National Science Foundation, J2)	("Chemical Abstracts", J43)

It seems to me that the advantages of the procedure
outweigh its disadvantages. Its main advantage is that it
makes it possible to process strings that are unpredictable.
Besides, it results in important reductions of the abstracts
under analysis, because of the great number of proper nouns
in those texts. The inclusion of certain items in list (16)
- in particular, comma, semi-colon, AT and THE - results
in drastic reductions which are worth mentioning here. Thus,
the inclusion of AT and THE permits the reduction to NPP
of the whole string "the Operations Research Group at the
Case Institute of Technology" (J2). The comma has been
included in the list because it can have a value similar
to AT; for instance, each of the following strings was also
reduced to NPP thanks to the inclusion of the comma in list
(16) : "the Library and Information Service, Paint Division,
Imperial Chemical Industries Limited" (J27), "the University
of Chicago, Graduate Library School" (J33). The inclusion
of AND may result in the lumping together (the reduction to
a single NPP) of two or more proper nouns, but I have found
that the amount of reduction achieved justified this lack
of delicacy : "the U.S. Patent Office and the National
Bureau of Standards" (J22). Similarly, the addition of
semi-colon to the list results in important reductions at
the expense of delicacy; thus, the following two strings
could both be reduced to a single NPP : "the International
Federation for Documentation; the Joint Engineering Council;
the German Documentation Society; and the Gmelin Institute
of Inorganic Chemistry and Related Sciences in the Max-Planck
Society for the Advancement of Science" (J34), "Bibliographic-
al; Reference and Documentation; International Exchange of
Publications Centre; and Technical Printing Division" (J39).

Only in one text has the procedure resulted in a clearly
wrong analysis, namely a wrong segmentation. The underlined
portion of the string "the National Institutes of Health
Russian scientific translation program" (J3) was reduced to
NPP because "National", "Institutes", and "Health" were
replaced by NOA CODED in the dictionary-lookup procedure.
Still, as "Russian" is replaced by a semantic code and
appears in the title ("Problems in communicating Russian
science") the information contained in "Russian" could be
preserved.

Loss of information is the price which needs to be paid
for the important reductions and simplifications achieved
by the procedure. I have found, however, that this price
is worth paying and that the reduction of most proper nouns
and capitalized expressions to NPP did not seriously affect
the indexing, even though it raises a problem. The problem
can be explained by returning to the example given above.
After application of the proper noun reduction procedure,
the beginning of abstract J3 has a coded form equivalent
to (18) :

(18)

> Problems in communicating Russian science.
> An outline of NPP scientific translation
> program (...) (J3).

For such strings, I have assumed that, when this is required
for indexing, the semantic status of NPP can be inferred from
the context, that is, established by some later rule in the
system. Thus in the above example, I have assumed that the
status of NPP as some institution (rather than some person
or some name of a machine or system)can be established on
the basis of the concatenation of the string "of NPP
scientific translation program" (in G.2 : section 2.4.3.)
and can be confirmed by the rest of the text. This is just
an assumption or working hypothesis. Compare the treatment

the string "the Rapid Selector program of the Bureau of
Ships" (J48) in which "Rapid Selector" was replaced by
semantic codes, while "the Bureau of Ships" was reduced to
NPP.

I do not wish to claim that all the assumptions I have
made are correct. There are other thinkable strategies then
the one I have adopted here : for instance, one could leave
the ambiguity of NPP unresolved. For lack of reliable knowledge
about the inference processes involved in examples like the
above, I prefer to leave the question open.

2.4.1. G.O

G.O refers to the first set of concatenation rules
operating immediately after the proper noun reduction procedure.
As already pointed out G.O works schematically as follows.
It concatenates only pairs of codes, but can operate in
several passes, one pass treating the rewriting obtained in
the preceding pass or passes as a single symbol. For instance,
B is concatenated with C, yielding the rewriting B+C, which
can then be concatenated with A to yield the rewriting
A+(B+C). From now on, I shall adopt the convention used in
this abstract example : I shall represent the result of a
first concatenation as in B+C and give the result of one or
more preceding concatenations between brackets, as in (B+C).
As far as possible, I shall represent the operations of G.O
and of the following 'grammars' by quoting the actual words
rather than the corresponding codes that were actually
processed or obtained by the machine in the experiment; these
codes would obscure rather than clarify my presentation,
which is centred on the ordering of sets of concatenation
rules, rather than on the codes on which the machine operates.
Unless otherwise specified, the rewritings resulting from the
concatenations preserve the information conveyed by the
concatenated symbols. When this is not the case, the word
corresponding to the information that has been preserved is
underlined in the lists given below. In this section, the
concatenations performed by G.O in the experimental corpus
are listed exhaustively.

The purpose of G.0 is to prepare the operation of the following 'grammars' by reducing certain strings of codes to a single code. The corresponding natural language strings are assumed to have the following characteristics :

(i) Semantically, they can be regarded as wholes, even though they appear graphically as multiple-symbol strings : for instance, 'more + Adjective' as compared to "better".

(ii) No material is insertable between their constituent symbols. Thus, "in order to" is concatenated at this stage, whereas "for the purpose of" is processed in G.2 because some adjectives such as "chief" can be inserted between "the" and "purpose".

(iii) As is perhaps more frequently the case, it is assumed that if insertions are possible, their occurrence could be tested, and the longer string could be processed by means of the same procedure and computer programme, or another such as that of G.2 : for instance, it is assumed that it is possible to add a rule to G.0 if one wishes to allow for the occurrence of "at the least", next to "at least". As already pointed out, G.0 and the following 'grammars' were set up just for the experimental corpus of 50 abstracts, and there is no doubt that various rules would have had to be changed of the experiment had been extended to other texts.

The other main assumptions underlying G.0 (and in fact also the other 'grammars') are the following :

(i) It is assumed that context-free concatenations of immediate constituents are justifiable in a corpus of abstracts like Janaske; for instance, it is assumed that a corpus like Janaske does not present the type of ambiguity occurring in a sentence like "The man who come in order to stop the fighting...." where at least

two possible segmentations depend on whether "order"
is analyzed as the main verb of the sentence, or whether
one reads "in order to"in the relative clause "who...".
The assumption proved to be correct in the experimental
corpus.

(ii) Should assumption (i) turn out to be incorrect, the
approach adopted will have the merit of showing to what
extent and in what specific ways assumption (i) is not
valid in a corpus of abstracts.

(iii) It is further assumed that unhappy consequences of
assumption (i) can be remedied by further rewritings
and code manipulations. A possible candidate for perform-
ing this task is an error-correcting subroutine or
'grammar'. The existence of such a subroutine has been
mentioned in a recent paper (Shapiro 1967).

The types of strings processed by G.O are listed under
the headings given below. The treatment of comma and semi-
colon is scattered under the various headings; I do not
know of any satisfactory treatment of this difficult problem
in the literature of computational linguistics. Except in a
few cases, the concatenations made in G.O are not directly
relatable to rules of Harris's string theory. Rather, they
serve to rewrite as a single code material which will later
be concatenated according to rules partly based on string
theory.

(i) Proper nouns.
Various proper nouns are concatenated at this stage
because they are equatable with one-symbol expressions :
compare, for instance, "United States" and "U.S.". The
proper nouns processed by G.O are listed below in (19):

(19)

- Federal+Government; the+(Federal+Government)(J5)
- United+States; the+(United+States)(J6, J30)
- Democratic+Republic; German+(Democratic+Republic);
  the+(German+Democratic+Republic)(J7)
- the+Americas(J29)
- Near+East; the+(Near+East)(J29)
- Federal+Republic;   (Federal+Republic)+of;
  (Federal+Republic+of)+Germany(J31)
- American+Documentation(J33)
- Soviet+Union(J42, J47)
- Rapid+Selector(J48)

(ii) Anaphoric or cataphoric expressions and pronouns.

The list of the strings brought under this heading is
given below in (20). Note the difference in meaning
between the occurrences of "the center" in J1 and J7.
Also note that, by mistake, various anaphoric expressions
have not been processed at this stage : "the classifica-
tion" (J19) "this system" (J26), "the Institute"(J39),
etc. Perhaps the most consistent and systematic treatment
of such the+N anaphoric strings would be at the stage of
G.2.

(20)

(a) - the+center(J1)
    - the+center (J7)
    - the+bureau(J48)

(b) - he+feels(J12)
    - viewing+it(J12)
    - making+it;(making   +it)+possible(J19)
    - he+explains(J20)
    - give+us(J38)
    - they+discuss(J38)
    - he+believes; (he+believes)+should(J40).

(iii) Genitives.

The genitives processed in G.0 are listed below in
(21) :

(21)
- (the+Federal+Government)+'s(J3)
- country+'s(J6)
- machine+'s(J9)
- NPP+'s(J15, J33)
- examiner+'s(J22)
- Hungary+'s(J45)
- (the+bureau)+'s(J48)

(iv) Words associated with Verbs.

The "that" introducing a clause after certain verbs has been concatenated with this verb; "to be","as" following certain verbs or cognate nominals have also been processed in this fashion :

(22)
(a) - concludes+that(J5)
    - concluded+that(J11)
    - recognizes+that(J12)
    - feels+that(J22)
    - appears+that(J22)
    - indicates+that(J26, J42)
    - (stated+emphatically)+that(J47;see.below)
(b) - recognized+as(J5)
    - (viewing+it)+as(J12)
    - (for+use)+as(J36, see below)
    - serve+as(J32)
(c) are+:(J38)
(d) to+be(J16, J22, J26, J35, J44)

(v) Wh-words.

Under this heading are grouped various expressions which can be associated with, or rewritten as, relative pronouns :

(23)
- who+is(J8)
- in+which(J9)
- that+which; of+(that+which)(J12)
- of+what(J16)
- in+which(J35)
- in+which; (the+ways)+(in+which); on+(the+ways+ in+which)(J40, see below).
- those+who; by+(those+who) (J40)
- ,+which(J46)
- in+which; ,+(in+which)(J48)

(vi) Conjunctions and prepositions.

The list is given below, in (24) :

(24)

(a) - only+ (rewritten as a conjunction:J7)
  - so+that(J16)
  - as+well;(as+well)+as(rewritten as a conjunction : J26, J46)
  - ,+since(J26)
  - not+merely;C2+(not+merely)(rewritten as a conjunction : J26 see below)
  - ;+and(several occurrences : rewritten as a conjunction)
  - ,+and(several occurrences : rewritten as a conjunction , C2 standing for "and")

(b) - in+order;(in+order)+to(J2, J45)
  - addition+to; in+(addition+to)(J5)
  - with+respect;(with+respect)+to(J11, J16)
  - ,+as;(,+as)+far; (,+as+far)+as (J11, see "are concerned", below)
  - according+to(J12, J38)
  - out+of(J16)
  - each+to(J22)
  - with+reference; (with+reference)+to(J33)
  - as+to(J40)
  - such+as; ,+(such+as) (J45, J50)
  - as+in (J47)
  - because+of(J49)

(c) - into+as; (into+as)+many(J21)
  - as+compared(J21)
  - by+:(J34)

(vii) Adjectives.

The types of strings concatenated and rewritten as adjectives are given below in (25) :

(25)

(a) - some+of(J6, J49)
  - each+of(J7)
  - all+of(J1, J19, J21)
  - certain+of(J13)
  - most+of(J13)
  - one+of(J19)
  - many+of(J21)
  - NUM+of(J24, J25, J32)

(b) such+a (J21, J28, J29)

(c) - large+scale; no+(large+scale)(J20)
    - long+range(rewritten as adjective : J29)
    - high+capacity(rewritten as adjective : J50)

(viii) Adverbs.

Various types of adverbs are concatenated with other
word-classes. Here, my rules correlate rather closely
with some of Harris's rules concerning the D (adverbs)
category. Thus, Harris treats the type given in (26) (a)
as left adjunct of P (preposition), the type of D given
in (26) (b) can be regarded as left adjunct or A
(adjective), the type given in (c) is analyzed by Harris
as right adjunct of V (verb), and the type given in (d)
is one of Harris's adjuncts of C (center string) (Harris
1965, 35-39).

                            (26)

(a) - particularly+among; ,+(particularly+among)
        (J5; rewritten as PADONG)
    - especially+in; ,+(especially+in)(J38, J46;
        rewritten as PIN)

(b) - the+most;(the+most)+economic(J11)
    - in+as; (in+as)+neat; (in+as+neat)+a (J12)
    - the+most; (the+most)+crit ... l(J16)
    - more+accurate(J22)
    - the+most; (the+most)+ ... rtant(J25)
    - well+ahead(J32)
    - manually+operated(J35)
    - more+precise(J36)
    - more+efficient(J42)
    - great+that; so+(great+that)(J43)
    - most+frequently(J48)

(c) - change+over(J1)
    - carried+out(J2)
    - points+out(J9, J38)
    - carried+on(J23, J39)

(d) - however+,;,+(however+,)(rewritten as NUL);
NUL+that(J12). See above : (he+feels)+that
- addition+,;in+(addition+,);,+(in+addition+
,)(J18, J22, J29 : the code for In+addition+
comma is amalgamated with that for period,
and only the code for period is retained)
- however+,;,+(however+,)(J32 : reduction to
the code for period, as above)

(e) at+least(J12)

(ix) Miscellaneous.

Under this heading, I have brought together various
concatenations which are mainly intended to bring about
preliminary simplifications for later stages in the process-
ing. Some of the concatenations performed at this stage
could have been more consistently performed at later
stages, as was later realized. One clear example of
inconsistency is the concatenation "mass+media"(J41)
where "mass" is a left adjunct of "media" which should
have been processed in G.2 (see section 2.4.3., below).
Another inconsistency is the reduction at this stage
(see v, above) of "on the ways in which" (J40) to a
single code equivalent to "on how", while a similar
string "the way in which" (J35) is only reduced to a
single code (XHOW) in G.2

At this stage, I also amalgamated "are" with "concerned"
and with a comma following in J11 in order to reduce the
string of codes corresponding to "the most economic, as
far as x are concerned," to three codes (ignoring x;
see above). The NUL code obtained after some of these
concatenations of G.0 was deleted by a special subroutine
operating immediately after G.0; this subroutine is
used again later, after G.1 (see next section) to delete
the ADV code for adverbs in cases where it has not been
deleted by concatenation rules. The two lists of
'miscellaneous' strings concatenated in G.0 are given
below in (27) and (28) :

(27)

- ,+etc.(rewritten as NUL);NUL+,(rewritten as NUL;
  J25)
- for+use(J26 : see "for+use+as" above)
- the+way(J35)
- ,+etc. (rewritten as NUL; J38)
- in+mind; kept(+in+mind)(J40)
- the+ways(J40)
- per+year(J43)
- of+view; point+(of+view)(J47)

(28)

- NUL+to; (NUL+to)+NUM; from+(NUM+to+NUM)(J13)
- March+NUM(J15)
- NUL+hours(two occurrences : J22)
- quarter+NUM(J33)
- of+between; (of+between)+NUM; (of+between+NUM)+
  and (rewritten as POF : J37)
- NUM+seconds(J37)
- NUM+to; (NUM+to)+one (rewritten as NUM : 2
  occurrences, J37)
- about+NUM; (about+NUM)+(per+year) (J43; see above)

2.4.2. G.1

As already noted, G.1 concatenates pairs of codes between which th.re is a code for conjunction. Schematically: AcB→ A+B. Moreover, in enumerations the rewriting(→) resulting from (a) precoding concatenation(s)can be concatenated with a conjoined code that follows : (A+B)cC→ A+B+C. The following (29) is an exhaustive list of the concatenations performed in G.1; they are given in the order of their first occurrence in the experimental corpus. Note that the treatment of conjunction in G.1 differs from that outlined by Harris who proposed to treat, say, conjunction+B as right adjunct of A (Harris 1965, 39-40). The treatment of A and B on the same footing (on either sides of the conjunction) which has been adopted here can be preferred on the grounds that : (i) in enumerations it is a bit odd to say that the second and following conjuncts are adjoined (added as something non-essential) to the first; (ii) various conjuncts are reversible without changing the meaning or affecting grammaticality (see, for instance, J25; on this, section 3.1.4).

My treatment of conjunction is a good example of the superficial nature of the processing performed by the system presented here. Consider the following sentence : "Lists service bureaus in Canada and the United States by province or state, and by city within province or state" (J30). Under one hypothesis of TG linguistics (see Part III and Stockwell et al. 1968), such a sentence would presumably be derived from a deep structure paraphrasable as : "(The author ?) lists service bureaus in Canada by province and he lists service bureaus in Canada by city within province, and he lists service bureaus in the United States by state and he lists service bureaus in the United States by city within state". The sentence as it appears in J30 would be

derived by so-called conjunction reduction. In the present
system, what can be regarded as already the result of a
transformational reduction is further reduced to codes
paraphrasable as : lists service bureaus in <u>an international
area</u> by <u>region</u>, and by city within <u>region</u>.

<div align="center">(29)</div>

- science+and+technology(J1)
- collecting+,+processing+,+storing+(,+and)
  +retrieving (rewritten as a code for '<u>documenting</u>' :
  J1)
- storage+and+retrieval (rewritten as a code for
  '<u>documentation</u>' : J1; J10, 2 occurrences; J11; J13;
  J15; J50)
- reasoning+and+<u>thinking</u>(J2)
- design+and+construction(J2)
- selection+,+form+,+distribution+,+acceptability
  (J3)
- medical+and+chemical(J4)
- interpreting+and+determining(J4)
- tracing+and+identifying(J4)
- research+and+development(J5, 2 occurrences)
- scientists+and+engineers(J5)
- scientific+and+technical(J6, 4 occurrences; J29)
- microfilm+and+photocopying(J7)
- problems+and+questions(J11)
- access+,+durability(J11)
- time+,+capacity+(,+and)+efficiency(J11)
- theorems+,+systems(J12)
- revised+and+enlarged(J13)
- research+,+development+(,+and)+testing(J14)
- documentation+and+librarianship(J14)
- individuals+or+organizations(J14)
- retrieval+and+reproduction(J14)
- indexing+,+cataloging+(,+and)+classification(J14)
- production+and+dissemination(J14)
- efficient+and+effective(J16)
- (of+what)+and+how(J16)
- discussion+and+comparison(J16)
- ideas+and+findings(J16)
- adequate+and+efficient(J16)
- methods+,+measures+and+results(J16)
- storage+or+retrieval(J16)
- criteria+and+procedures(J17)
- requirements+,+criteria+(,+and)+measures(J17)
- evaluate+and+compare(J17)
- existing+and+proposed(J17)
- statistical+,+syntactical+and+interrelational(J18)
- concepts+and+interrelationships(J19)

- concepts+and+interrelations(J19)
- (NUM+hours)+or+less(J22)
- input+and+output(J23)
- archives +,+libraries(J25)
- libraries+,+archives(J25)
- German+,+French+,+Italian(J25)
- institutions+,+subjects+(,+and)+personnel(J25)
- code+and+card(J26)
- edge-punched+,+hand-sorted(J26)
- reports+and+memoranda(J28)
- Europe+,+(the+Near+East)+and+Africa(J29)
- Asia+and+Australasia(J29)
- scientific+or+technical(J29)
- Canada+and+(the+United+States)(J30)
- province+or+state(J30 : 2 occurrences)
- Canada+,+France+,+Holland+,+(Federal+Republic+ of+Germany)+,+U.S.A.+(,+and)+Italy(J31)
- fast+,+large(J32)
- needs+and+possibilities(J34)
- selection+and+reproduction(J35)   .
- microfilming+,+punched-cards+(,+and)+xerography (J35)
- manually+or+mechanically(J35)
- filing+and+reproduction(J35)
- recorded+and+filed(J35)
- small+,+(manually+operated)+,+portable(J35)
- design+and+construction(J37)
- display+and+reproduce(J 37)
- NUM+and+NUM(J37)(J43)
- role+and+significance(J38)
- origins+and+development(J38)
- science+and+technical(J38)
- card-indexes+,+libraries(J38)
- source+,+content+and+value(J38)
- graphical+,+pictorial+,+auditory+,+visual(J38)
- collecting+and+investigating(J38)
- sources+and+use(J38)
- Rumanian+,+Hungarian+,+Polish+,+Serbo-croatian+ and+Spanish(J38)
- practices+and+attitudes(J40)
- acquisition+and+handling(J41)
- decentralized+and+inefficient(J42)
- means+and+personnel(J42)
- past+,+present+(,+and)+future(J43)
- U.S.+and+U.S.S.R.(J43)
- abstracting+and+indexing(J44)
- librarianship+and+documentation(J45)
- documentation+and+economy(J47)
- professional+and+organizational(J47)

- whole-page+or+total-document(J48)
- lines+and+paragraphs(J48)
- requirements+and+advantages(J49)
- timely+and+accurate(J50)
- illustrations+and+prices(J50)
- manufactured+or+sold(J50)
- obsolete+and+discontinued(J50)
- classified+and+indexed(J50)

In the above examples I have found only one clear case of unresolvable ambiguity. It is quoted below in (30) (a) as it appears in its context; the bracketings in (b) represent the interpretation I have arbitrarily chosen, and those in (c) represent another possible interpretation. The ambiguity, however, does not affect the indexing by means of the headings of Gardin-Lévy. In (31), I give what I believe to be the only wrong analysis (b) which has crept into the experiment with G.O. The mistake, however, does not affect the indexing either, and what I believe to be the correct analysis (c) could have been obtained by later rules in the system : only "sources and use" should be concatenated at this stage, and the other conjunction reduced in G.6.

(30)

(a) interpreting and determining the content of foreign prescriptions (J4)

(b) (interpreting and determining) the content of foreign prescriptions.

(c) interpreting and (determining the content of) foreign prescriptions.

(31)

(a) collecting and investigating the sources and use of documents (J38)

(b) (collecting and investigating) the (sources and use) of documents.

(c) collecting and (investigating the sources and use of) documents.

Finally, let us note that after G.1 some codes for adverbs are deleted by a special subroutine. Specifically, the subroutine deletes ADV (as for "thereby" in J9, and "at+least" in J12), and ADX not deleted by ▄ some earlier rules, as was "only" (J9). However, the semantic codes replacing some adverbs are not deleted : for instance, "logically" (J12), "either (=ADV : deleted) (manually+ or+mechanically)" (J35). The adverbs "then" and "very" have (perhaps mistakenly) been given special codes (XTHEN, XVERY), which are not deleted by the subroutine.

:

2.4.3. G.2

G.2 concatenates in succession strings of six, five, four, three, and two codes into a single code. The strings of six codes had to be processed by a special programme, because the 105 characters necessary for six codes plus a code for the rewriting (15x7) exceeded the 80 characters provided for in the programmes set up for the processing of shorter strings, and available on the punched-cards serving as input. The special programme had been written to process strings of up to 7 codes, but this appeared to be unnecessary, since no rule of that length was needed in the experimental corpus.

There is an obvious reason for beginning by recognizing and reducing the longer strings and for ending with the shorter ones. Thus, the following string of 6 codes given in (32) (a) needs to be processed first, since the other strings given in (32) can be regarded as subsets of (a) :

(32)

(a) <u>6 codes</u> : of an international technical information system (J29)

(b) <u>5 codes</u> : - of a technical information system
- of an international information system

(c) <u>4 codes</u> : - of an information system
- of an international system

(d) <u>3 codes</u> : of a system.

Grammatically, the strings processed by G.2 fall into the types which are given below. In this section, I shall just list and briefly illustrate the types of strings in question, because a complete list of all the strings processed in G.2 would amount to quoting practically the whole experimental corpus. Besides, such a list is unnecessary since many examples of strings processed in G.2 will have to be

given in any case in order to explain the following
'grammars'. Only the examples of six-code strings are
listed exhaustively. The concatenation of prepositions with
material following has been performed at this stage (cp.
Bély et al. 1970, 99 : "groupe prépositionnel"). By Noun
in (ii) are meant not only nouns in the strict sense but
also forms that are relatable to verbs but have a grammatical
behaviour similar to nouns, in particular as far as their co-
occurrence with articles and with prepositions is concerned.

The only grammatically incorrect consequence of the
treatment of prepositions adopted concerns what Harris would
presumably analyze as Nof left adjuncts of Nouns (Harris
1965, 36). In my experiment, the strings of (33) (a) have
been segmented as shown in (b), whereas they should have
been segmented as in (c) where both the genitives and what
is underlined in (a) are analyzed as left adjuncts of N :

(33)

(a) a general review of the scope of the Federal
Government's role in ... and a review of the
character of the Federal Government's scientific
information activities (J5)

(b) (a general review) (of the scope)(of the
Federal Government's role)(in...) and (a review)
(of the character)(of the Federal Government's
scientific information activities)

(c) (a general review)of((the scope of)(the
Federal Government's)role) (in...)and (a review)
of ((the character of)(the Federal Government's)
scientific information activities).

The only justification for the treatment adopted here
(compare "some+of", etc., in G.0) is that, by lumping
together graphically similar strings like "of the scope"
and "of the Federal Government's role", one pass in the
processing can be saved. Similarly, what Harris treats as
'nominalized' center strings is here broken down into two
strings. Thus, about the string "his description of the
experiment", Harris says in essence that "description" is

a nominalization of "describe", that "of the experiment"
functions as Object, and that the whole string can function
as Subject or Object of a center string (Harris 1965, 29-31);
here, it would be broken down into "his description" and
"of the experiment".

Lastly, note that some concatenations performed in
G.O (ii) such as "he+feels" (J12) could have been more
consistently grouped together with some exemplified below
in (i) (c), such as "this+reviews" (J2) (see my discussion
in Chapter 3.2.).

(i) (a) Auxiliary+Verb and be+x.

Harris (1965, 37 and 32) treats auxiliaries as left
adjuncts of V, be as a Verb and various items following
be (what is here represented as x) as types of Objects :
Also note that I have treated 'constitute' like 'be'.

EXAMPLES

4 codes : - can+be+a+success(J20)
- is+an+organizing+system(J21)
- will+be+(more+accurate)+than(J22)
- constitutes+a+major+change(J26)
- are+NUM+other+ares(J32)
- being+a+(small+,+manually+operated+,+
portable)+unit(J35)
- are+the+retrieval+units(J48)

3 codes : - must+be+(recognized+as)(J5)
- is+magnetic+tape(J11)
- will+be+lost(J21)
- can+be+punched(J26)
- will+make+possible(J32)
- is+(to+be)+reproduced(J35)
- is+then+printed(J35)
- has+been+designed(J37)
- should+be+replaced(J42)
- constitute+other+centers(J46)
- are+being+intensified(J49)

2 codes : - was+planning(J1)
- is+variable(J19)
- are+fixed(J19)
- is+dev .1oped(J21)
- are+gratifying(J22)
- will+require(J24)
- is+   +beginning(J24 : the blank
  indicates the position of an adverb
  deleted after G.1)
- will+(serve+as)(J32)
- will+make(J32)
- is+   +(well+ahead)(J32)
- is+   +suited(J35)
- are+available(J35)
- (are+:)+(graphical+,+pictorial+,+
  auditory+,+visual)(J30)
- is+(carried+on)(J39)
- would+permit(J42)
- were+NUM(J43)
- is+(so+great+that)(J43)
- continue+unchanged(J43)
-       +being+undertaken(J45)
- is+   +possible(J47)

(b) wh-word+(a) (cp. Harris, loc. cit. : wh-.   ,
are treated as a part of right adjuncts of :   also
whiz deletion in TG linguistics : for insta.. :,
Langendoen 1970)

4 codes : - which+has+been+developed(J5)
- which+may+be+ordered (J12)
- (of+that+which)+may+be+written(J12)
- which+would+be+based(J29)
- which+(he+believes+should)+be+(kept+
  in+mind)(J40)
- which+is+not+reported(J41)

3 codes : - which+was+established(J1)
- that+are+responsible(J2)
- that+will+determine(J12)
- which+would+permit(J26)
- (,+which)+is+augmented(J16)

(c) Cp. (a) and (b)

EXAMPLES

4 codes : - (who+is)+   a+subject+specialist (J8)
- it+can+be+   +modified(J26)

3 codes : - it+is+   +necessary(J8)
- it+is+(concluded+that)(J11)
- has+broad+applications(J13)
- (to+be)+   +(the+most+critical)+
  problem(J16)

- it+can+handle(J20)
- it+is+proving(J22)
- there+would+be(J29)
- had+its+origin(J33)
- was+so+present(J34)
- is+(to+be)+reprcduced(J35)
- it+is+not(J36)
- is+purported+(to+be)(J44)
- it+is+(stated+that)(J47)
- it+is+(classified+and+indexed)(J50)

2 codes : - this+reviews(J2 : 'reviews' was coded
as a homograph; in the rewriting of
G.3, it is represented as a verb)
- which+uses(J9)(J48)
- it+   +possesses(J11)
- this+is(J13)(J18)(J23)(J26)(J44)
- NPP+considers(J16)
- is+that(J19)
- there+are(J21)
-        +it+(appears+that)(J22)
- (to+be)+held(J35)
- it+works(J36)
- it+is(J36)
- which+stores(J37)
- one+stores(J37)
- lays+stress(J38)
- there+is(J38)
- (by+those+who)+manage(J40)
- who+observed(J44)

(ii) (a) Left adjunct(s)+Noun (cp. Harris 1965, 36)

Note that the concatenations illustrated below ignore
structural differences in the relations between the
constituents of the strings : the working of the computer
programme are based only on the number of codes to
be recognized as forming a string. Also note the processing
of some semi-colons at this stage.

EXAMPLES

6 codes(only example) :
(the+bureau+'s)+well-established+NPP+
information+retrieval+program(J48)

5 codes : - this+(country+'s)+(scientific+and+
technical)+information+ills(J6)
- the+Automatic+Microfilm+Information
System(J36)
- the+NULL+   +centralized+information
+service(J44)

4 codes : - the+botanical+drug+literature(J4)
- improved+information+(storage+and+retrieval)+systems(J15)
- scientific+information+retrieval+systems(J17)
- (no+large+scale)+information+retrieval+system (J20)
- NUM+established+techniques+:(J35)
- a+very+useful+supplement(J38)
- the+Yugoslav+Bibliographical+Institute (J39)
- the+technical+industrial+libraries (J46)
- (timely+and+accurate)+factual+comparable+data(J50)

1 codes : - a+general+description(J1)
- the+initial+(reasoning+and+thinking)(J2)
- partial+contents+:(J15)
- (one+of)+its+features(J19)
- the+(examiner+'s)+time(J22)
- a+fast+system(J36)
- its+present+organization(J38)
- other+trade+publications(J41)
- the+   +increasing+need(J42)
- the+same+year(J43)
- the+(Rapid+Selector)+program(J48)

2 codes : - its+activities(J1)
- user+requirements(J14)
- theoretical+studies(J14)
- documentation+centers(J25)
- user+information(J25)
- (such+as)+those(J28)
- much+work(J32)
- (fast+,+large)+memories(J32)
- the+simplest(J35)
- (the+way)+(in+which)(J35; rewritten as XHOW)
- a+system(J36)
- the+second(J37)
- the+(sources+and+use)(J38)
- science+libraries(J40)
- existing+tools(J43)
- its+core(J45)

(b) <u>Proposition+(s)</u> (see also examples with semi-colon).
This includes simple <u>PN</u> adjuncts of center strings
(Harris 1965, 39) such as "in the second chapter" (J38).

<div align="center">EXA..PLES</div>

<u>6 codes</u> (exhaustive list)

- in+the+following+NUM.+areas+: (J14)
- to+some+other+search+file+system(J21)
- of+an+international+technical+information
  +system(J29)
- (with+reference+to)+(NPP+'s)+NUM+basic
  +building+rules(J33)
- of+the+Automatic+Microfilm+Information
  +System(J36)
- on+a+cathode+ray+tube+screen(J37)
- of+the+centralized+science+information+
  services(J44)

<u>5 codes</u> : - to+a+mechanized+(storage+and+retrieval)
  +system(J1)
- of+NPP+scientific+translation+program
  (J3)
- <u>into</u>+the+following+parts+: (J6)
- from+(certain+of)+their+basic+charac-
  teristics(J12)
- of+scientific+information+retrieval+
  systems(J17)
- of+the+NUM+element+, (J24)
- to+(NUM+of)+(the+most+important)+
  Swiss+(libraries+,+archives)(J25)
- in+the+pertinent+geographical+region(J25)
- (for+use+as)+a+detailed+search+code(J26)
- of+a+new+subject+index(J27)
- in+a+standard+eighty-column+punched-
  card(J35)
- in+the+traditional+(mass+media)+effort
  (J41)
- of+the+large+production+centers(J46)
- for+all+known+microreproduction+equipment
  (J50)

<u>4 codes</u> : - in+the+(medical+and+chemical) literature
  (J4)
- at+(NPP+'s)+NUM+NPP(J15)
- into+NUM+classes+: (J16)
- in+this+system+, (J19)
- between+the+(input+and+output)+
  languages(J23)
- :+a+pragmatic+approach(J27)
- of+the+various+types(J27)
- :+one+for+(the+Americas)(J29)

- by+NUM.+appendices+i(J33)
- over+a+special+aperture(J35)
- having+linear+reduction+ratios(J37)
- in+the+NUM+chapter(J38)
- by+NUM+divisions+:(J39)
- (as+to)+certain+general+(practices+ and+attitudes)(J40)
- for+a+(more+efficient)+organization(J41)
- of+scientific-technical+information+ centers(J46)
- from+the+(professional+and+organizational) +(point+of+view)(J47)
- whatever+the+financial+cost(J47)
- by+the+proposed+NPP(J47)
- on+a+(whole-page+or+total-document)+ basis(J48)
- to+many+new+areas(J50)

3 codes :
- for+the+purpose(J1)
- for+the+dissemination(J2)
- of+the+scope(J5)(J42)
- (in+as+neat+a)+fashion+as(J12)
- (with+respect+to)+the+latter(J16 : rewritten as NUL, and deleted after G.2).
- as+a+basis(J20)
- (into+as+many)+sections+as(J21)

- in+the+past(J27)
- :+(long+range)+planning(J29)
- off+the+real(J35)
- before+being+installed(J37)
- on+the+(role+and+significance)(J38)
- PIN+(science+and+technical)+subjects (J38)
- (according+to)+the+(sources+,+content +and+value)(J38)
- of+the+Institute(J39)
- :+a+survey(J40)
- in+NUM+journals (J43 : wrong analysis)
- in+the+world(J44)
- of+(Hungary+'s)+network(J45)
- of+scientific-technical+information(J46)
- by+a+NPP (J46)
- by+   +qualified+personnel(J47)
- on+old+ones(J49)
- (because+of)+their+applications(J49)
- (such+as +(high+capacity)+(storage+ and+retrieval)(J50)

2 codes : - of+NPP(J1)
- (in+order+to)+achieve(J2)
- of+(selection+,+form+,+distribution+,
  +acceptability)(J3)
- (as+far+as)+costs(J11)
- as+   +appropriate(J12 : this 'as'
  should have been concatenated like
  others in G.O)
- ,+NUM(=,1957-1958 : J14)
- :+(statistical+,+syntactical+and+
  interrelational)(J18)
- in+that(J22)
- (to+be)+possible(J22)
- to+   +handle(J22)
- (each+to)+(NUM+hours+or+less)(J22)
- :+(German+,+French+,+Italian)(J25)
- (to+be)+used(J26)
- a+minute(J35)
- in+Rumania(J38)
- in+NUM(J38)
- :+(past+,+present+,+and+future)(J43)
- for+developing(J43)
- (,+such+as)+decentralization(J45)
- (in+order+to)+build(J45)
- PIN+Czechoslovakia(J46)
- (as+in)+(the+Soviet+Union)(J47)
- nor(=XNOR)+desirable(J47)

2.4.4. G.3

Like G.1, G.3 concatenates conjoined strings, and it uses the same computer programmes as G.1. Since the strings processed at this stage are by far fewer in number than those processed in G.2, I shall give a complete list of these conjoined strings below in (34). As will be noted, many concatenations performed at this stage operate on the results of concatenations performed in G.2 : for instance "its+activities"(J1).

When some point deserves to be made about the rewriting, the quotation is followed by a paraphrase of, and a comment on, this rewriting. The paraphrase is given between single quotation marks ('——'). No current theory, as far as I can see, is able to provide a basis for my rewritings. My paraphrases and my comments must therefore be regarded as statements of semantic problems raised by my experiment. Many of these problems have to do with what I have referred to earlier as the inference-making process involved in indexing, and in the resulting loss of information. Because of these problems, I shall henceforth give up the convention which consisted of underlining the string whose information is preserved in the rewriting.

G.3 processes strings of the following types :
(i) Conjoined nouns with or without left adjuncts : for instance, "NPP+and+(its+activities)"(J1)
(ii) Conjoined strings consisting of Preposition+left djunct(s)+Noun preceding, and left adjunct(s)+Noun following the conjunction; these strings are transformationally relatable to type (iii) : for instance, "(in+BDP)+and+(information+systems)"(J15).

(iii) Conjunctions of <u>Preposition+left adjunct(s)+Noun</u> :
"<u>of</u> drawing-office supplies and <u>of</u> reproduction drawings"
(J35).

(iv) Conjunctions of verbs.

(v) Strings relatable to the former by virtue of preceding
concatenations.

(vi) Various strings separated by a comma : see J26.

<div align="center">(34)</div>

- NPP+and+(its+activities)(J1); rewritten as a code
  paraphrasable as noun singular, definite, insti-
  tution; I have assumed that "its activities"
  imposes this interpretation on the NPP that
  precedes.

- (of+selection+,+form+,+distribution+,+accepta-
  bility)+(,+and)+(translation+rights)(J2);
  rewritten as a code paraphrasable as 'of the
  development of translations'; the discussion on
  asymmetric conjunction in Part III is devoted
  to what happens in this type of conjunction.

- (the+Federal+Government)+and+(U.S.+scientific+
  information)(J5); rewritten as the code already
  obtained for the second conjunct, namely
  'national documentation', which corresponds to
  heading 5 of the Gardin-Lévy concordance (see
  Appendices 1 and 2).

- (of+microfilm+and+photocopying+services)+or+
  (photocopying+service)(J7); rewritten as 'of
  photocopying services'.

- (of+a+scientific+information+officer)+and+
  (a+librarian)(J8); rewritten as 'of a documenta-
  list'.

- (among+magnetic+drums)+,+(magnetic+tape)+,
  +(magnetic+wire)+,+(ferromagnetic+cores)+,+
  punched-cards+,+(punched+tapes)+(,+and)+(photo-
  graphic+tapes)(J11); rewritten as 'among, plural,
  storage equipment'.

- (with+respect+to+their+access+,+durability)+,
  +(storage+and+retrieval+time)+,+(capacity+,+
  and+efficiency)(J11); rewritten as 'with respect
  to qualities'.

- (in+EDP)+and+(information+systems)(J15); rewritten
  as 'in data processing and documentation systems'.

- (in+information+retrieval)+and+(data+acquisition)
  (J15); rewritten as 'in documentation'.
- (of+information+retrieval+systems)+;+(improved+
  information+storage+and+retrieval+systems)(J15);
  rewritten as 'of documentation systems'.
- (are+many)+,+(including)(J21); rewritten as
  'include'.
- (steroid+compounds)+and+NPP(J22); rewritten as
  'topics in a particular discipline', on the basis
  of "<u>steroid compounds</u>"; here, NPP is actually
  given a wrong interpretation; the text reads
  '<u>ILAS</u>' which stands for "<u>Interrelated Logic
  Accumulating Scanner</u>", a "punched-card machine"
  mentioned in J322.
- (archives+,+libraries)+and+(documentation+
  centers)(J25); rewritten as 'documentation
  institutions'.
- (to+NUM+of+the+most+important+Swiss+libraries+,
  +archives)+,+(documentation+centers)(J25);
  rewritten as 'to national documentation institu-
  tions', a rewriting already chosen for the first
  conjunct in G.2.
- (historical+development)+,+holdings+,+(special+
  collections)(J25); rewritten as 'history and
  holdings'.
- (all+functional+units)+(,+and+not+merely)+(the+
  filing+categories)+,+(can+be+punched)(J26);
  rewritten as 'units and classification entries
  can be punched'; I have found no better way of
  treating the comma preceding the verb; I have not
  been able to figure out what is meant by "func-
  tional units".
- (a+search+code)+(as+well+as)+(a+filing+order+
  code)(J26); rewritten as 'search and storage
  code'.
- (one+for+the+Americas)+;+(one+for+Europe+,+the
  +Near+East+and+Africa)+;+(one+for+Asia+and+
  Australasia)(J29); rewritten as 'for name of
  international area'.
- (by+province+or+state)+(,+and)+(by+city)(J30);
  rewritten as 'by geographical region'.

- (of+drawing-office+supplies)+and+(of+reproduction
  +drawings)(J35); rewritten as 'of equipment and
  reproductions'.
- (NUM+chapters)+and+(a+subject+index)(J38);
  rewritten as 'parts of document'.
- abroad+and+(in+Rumania)(J38); rewritten as 'in
  an international area'.
- (research+work)+and+planning(J3C); rewritten as
  'research'.
- (house+organs)+and+(trade+publications)(J41);
  rewritten as 'special documents'.
- (of+house+organs)+and+(other+trade+publications)
  (J41); rewritten as 'of special documents'.
- history+and+(further+development)(J46); rewritten
  as 'history'.
- NPP+and+(the+documentation+branches)(J46);
  rewritten as 'the parts of documentation institu-
  tions'.
- (of+the+research+institutes)+or+(of+the+large+
  production+centers)(J46); rewritten as 'of the
  institutions'.

2.4.5. G.4

G.4 uses the same computer programmes as G.0. It
concatenates only two codes at a time, but it can concatenate
a code obtained by one rewriting with a code following, and
it can also operate in several passes : see the examples of
J18, J19 and J22 given below in (35). Since the strings
processed in G.4 are relatively few in number, I give a
complete list of these strings below in (35). G.4 concatenates
the following types of strings :

(i) Nouns which can be partly analyzed as left adjuncts of
Nouns, together with what has been concatenated with
these left adjuncts (including prepositions) in G.2,
are concatenated with the Noun they are adjoined to and
what has been concatenated with this Noun. Consider,
for example, the string "of the character of the Federal
Government's scientific information activities" (J5).
I assume that according to Harris's string theory, this
string would be represented by a string formula para-
phrasable as "P(preposition of)+left adjuncts ofN+N
(=activities)" (Harris 1965, 36 on left adjuncts of N :
"group effort", "wild plan", "a sort of investigator").
As pointed out in section 2.4.3., however, the segmenta-
tion on which G.2 is based is somewhat different, and can
be represented as "of the character" and "of the Federal
Government's scientific information activities",
respectively. G.4, then, concatenates the codes for
these two types of strings : (of the character)+(of the
Federal Government's scientific information activities).
In the rewriting, the information corresponding to what
Harris would analyze as an Nof left adjunct of N ("a sort
of investigator", "the character of...activities") is
done away with. When this Nof adjunct is preceded by a
preposition ("of the character") this preposition is

preserved in the rewriting : thus, the rewriting assigned to the above example from J5 is paraphrasable as 'of national governmental documentation'. It is the omissibility of "the character of" which served me as a basis for deciding that a given Noun can be analyzed as left adjunct of a Noun. Thus, Harris would say that the excision of "a sort of" preserves sentencehood in "He is a sort of investigator" (Harris, loc. cit.). Similarly, the excision of "the character of" in "a review of the character of the Federal Government's scientific information activities" (J5) results in a grammatical string (Note, however, that the excision of "investigator" in Harris's example, and of "activities" in the extract from J5 have different effects; some of the special properties of the words "information" and "activities" in the above example will be discussed in Part III).

(ii) The theoretical status of the second type of string processed at this stage is unclear in string theory and is disputed in other theories, as far as I can see. An example is "for the purpose of" (J1). In G.4, a string such as "for the purpose" is concatenated with the "of Ving" that follows : (for the purpose)+(of contributing)(J1). The rewriting is paraphrasable as "for contributing". My justification for this type of concatenation lies precisely in this paraphrasability, and in the analogy between what happens here and what happens when an Nof adjunct of N is left out.

(iii) Like that just mentioned, the third type of concatenation performed in G.4 is based on a paraphrase relation, since its theoretical status is unclear or disputed. For example, "presents" is concatenated with the following "a review" (J5) and the rewriting is paraphrasable as "verb, reviews".

(iv) Some special cases : see examples below in J6, J7, J8,
    J19, and similar examples in other abstracts. These
    include the rewriting of "articulate" sentences ("<u>I thank
    you</u>") as "semi-articulate" ("<u>Thank you!</u>") or "inarticulate"
    sentences ("<u>Thanks!</u>") (Jespersen 1965, 308; see Chapter
    3.2). Also note the treatment of codes corresponding to
    strings with semi-colon, as in J6. An example of a drastic
    simplification <u>without</u> grammatical motivation is the
    deletion by rewriting of the codes for "One of its
    features is that" in J19. Compare the treatment of "The
    results are gratifying in that" (J22).

    Like in the preceding section, the paraphrases of the
rewritings given below in (35) after the examples are intended
to point to various problems which I have experienced in
setting up the codes for my experiment, particularly in
connection with the loss of information which takes place'
during the process of indexing.

                          (35)
        - (for+the+purpose)+(of+contributing)(J1);
          rewritten as 'for contributing'; more precisely,
          "for the purpose of" was rewritten as a code
          standing for 'Preposition of purpose'.

        - (in+order+to+achieve)+(a+net+improvement)(J2);
          rewritten as 'for improving'; more precisely,
          the code for the rewriting adopted in the experi-
          ment is paraphrasable as 'Preposition of purpose,
          Verb with some qualitative sense'; this loss of
          information corresponds to the fact that the
          specific meaning of 'for improving' need not be
          retained to index the abstract in which the
          string in question occurs. Note a similar loss
          of information in connection with "<u>net</u>".

        - (in+addition+to+presenting)+(a+general+review)
          (J5); rewritten as 'in addition to reviewing'.

        - (of+the+scope)+(of+the+Federal+Government+'s+
          role)(J5); rewritten as 'of governmental role'.

- (of+the+character)+(of+the+Federal+Government+
's+scientific+information+activities)(J5);
rewritten as 'of national governmental documen-
tation'.

- presents+(a+review)(J5); rewritten as 'Verb,
reviews'.

- (an+integral+part)+(of+research+and+development)
(J5); rewritten as 'research'.

- (into+the+following+parts+:)+(the+present+situa-
tion)(J6); rewritten as 'into the situation'.

- (the+address)+(is+given)(J7); rewritten as 'Gives
the address' (see 3.2 on the relations between
passive and semi-articulate sentences).

- (additional+information)+concerns (J7); rewritten
as 'verb, studies (or : discusses)' (see 3.2.).

- coverage+provided (J7); rewritten as 'coverage'.

- (a+syllabus)+(is+presented)(J8); rewritten as
'presents a syllabus'; compare "the address is
given" in J7.

- (for+a+student)+(who+is+a+subject+specialist)
(J8); rewritten as 'for a student'.

- (a+comparison)+(is+made)(J11); rewritten as
'a comparison'; this rewriting has the effect
of converting an articulate sentence into a semi-
articulate one (sec 3.2.).

- (NUM+suggestions)+(are+advanced)(J11); rewritten
as 'suggestions' (Cp. preceding example).

- (of+the+entire+body)+(of+written+knowledge)(J12)
rewritten as 'of information'.

- (a+finite+series)+(of+conjoined+propositions)(J12);
rewritten as 'topics of a science related to
documentation' (Group 4 of Gardin-Lévy : see
Appendix 1).

- (binary-octal+designations)+(are+established)(J12);
rewritten as 'Develops codes' (see 3.2.).

- (from+the+field)+(of+logic) (J12); rewritten as
'from a science related to documentation' (see
"conjoined propositions" in J12 above).

- (substantial+proportion)+(of+that+which+may+
be+written) (J12); rewritten as 'what may be
written'.

- (this+is)+(a+revised+and+enlarged+edition)(J13);
rewritten as 'edition' (see 3.2. on inarticulate
sentences).

- (most+of+the+material)+(has+broad+applications)
  (J13); rewritten as 'can be used',a rewriting
  already assigned to "has broad applications"
  in G.2.

- (in+the+following+NUM+areas+:)+coding (J14);
  rewritten as 'in coding'.

- (partial+contents+:)+advances (J15); rewritten
  as 'includes advances'.

- (over+a+span)+(of+NUM+years) (J16); rewritten
  as 'over years'.

- (into+NUM+classes+:)+need (J16); rewritten as
  'into need'.

- advantages+(of+the+new+techniques)(J16); rewritten
  as 'the techniques', a rewriting already assigned
  to "(of) the new techniques" in G.2.

- recommendations+(were+made) (J17); rewritten as
  'recommendations'.

- (this+is)+(a+description)(J18); rewritten as
  'a study' (see 3.2.).

- (of+the+following+types)+(of+information+retrieval
  +systems)(J18); rewritten as 'of documentation
  systems+colon' ; this rewriting in turn is conca-
  tenated with the following code for (:+statistical
  +,+syntactical+and+interrelational)and the
  rewriting of the pair is paraphrasable as 'of
  documentary languages'.

- (NUM+types)+(of+punched-cards)(J19); rewritten
  as 'punched-cards'.

- (one+of+its+features)+(is+that)(J19); the
  rewriting obtained is itself concatenated with
  a code for period that precedes, which is preserved
  as the final rewriting; in other words, the string
  'period+one of its features is that' is reduced
  to the code for period. Compare the deletion of
  the code for various adverbs : a discourse
  connective like "In particular" would also have
  been deleted if it had occurred instead of "One
  of its features is that".

- (it+can+handle)+searching(J20); rewritten as
  'it can search'.

- (there+are)+(different+codes)(J21); rewritten as
  'Noun, codes'.

- (the+advantages)+(of+this+inverted+search+file)
  (J21); rewritten as 'inverted file'.

- (a+systematic+procedure)+(is+developed)(J21); rewritten as 'develops   procedure'.

- (the+results)+(are gratifying)(J22); the rewriting of the two codes is itself concatenated with the code for the following (in+that); the second rewriting is then concatenated with the code for a period that precedes, which is preserved as the final rewriting; these succes-sive rewritings, then, have the effect of replacing the codes for ".The results are gratifying in that" by the code for period.

- (from+an+average)+(of+NUM+hours)(J22); rewritten as 'from hours'.

- (this+is)+(an+explanation)(J23); rewritten as 'a study'.

- (one+basic+part)+(of+their+program)(J23); rewritten as 'their program'.

- work+(is+   +beginning)(J24); rewritten as 'work'.

- (of+the+NUM+element+,)+(machine+languages)(J24); rewritten as 'of machine languages'.

- (descriptive+data)+include (J25); rewritten as 'Verb, studies'.

- (this+is)+(a+modification)(J26); rewritten as 'a modification'.

- (information+retrieval)+(:+a+pragmatic+approach)(J27); rewritten as 'a documentation method".

- includes+(a+short+account)(J27); rewritten as 'Verb+study'.

- (of+the+various+types)+(of+subject+indexes)(J27); rewritten as 'of documentary languages', a rewriting already assigned to "of subject indexes" in G.2.

- includes+discussion(J29); rewritten as 'verb+ study'.

- including+description (J31); rewritten as 'Ving+ study'.

- provide+(automatic+reference)(J32); rewritten as 'Verb+mechanical documentation'.

- (higher+densities)+(of+storage)(J32); rewritten as 'storage'.

- (to+the+deluge)+(of+technical+literature)(J32); rewritten as 'to the literature'.

192.

- (of+the+principal+features)+(of+the+classifica-
tion)(J33); rewritten as 'of the documentary
language'.

- (by+NUM+appendices+:)+(classified+guide)(J33);
rewritten as 'by a guide'.

- (by+:)+NPP(J34); rewritten as 'by NPP'.

- (NUM+established+techniques+:)+(microfilming+,
+punched-cards+,+and+xerography)(J35); rewritten
as 'microfilming+punched-cards+xerography'.

- (at+a+speed)+(of+NUM+feet)(J35); rewritten as
'at quantitative information', a rewriting already
assigned to '(of) NUM feet' in G.2.

- (in+a+space)+(of+NUM+square+feet)(J35); rewritten
as 'in quantitative information', a rewriting
already assigned to '(of) NUM square feet' in G.2.

- details+(are+given)(J35); rewritten as 'study'.

- (with+figures)+(of+savings)(J35); rewritten as
'with quantitative information', a rewriting
already assigned to '(of) savings' in G.2.

- (it+is)+(a+fast+system)(J36); rewritten as
'a system'.

- (it+is+not)+(a+system)(J36); rewritten as 'not
a system'.

- (of+elements)+(of+a+mechanized+library+system)
(J37); rewritten as 'of a mechanized documentation
system'.

- (on+feasibility+models)+(of+a+memory)(J37);
rewritten as 'on future storage equipment'.

- (in+the+form)+(of+microphotographs)(J37);
rewritten as 'as microphotographs'.

- (are+undergoing)+(final+tests)(J37); rewritten
as 'are being tested'.

- (an+attempt)+(is+made)(J38); rewritten as
'an attempt'.

- (the+different+types)+(of+documentation)(J38);
rewritten as 'documentation, comparison'; this
rewriting is concatenated with the code for
(points+out) that precedes, and the pair of codes
is rewritten as 'compares+documentation'.

- (there+is)+(a+very+useful+supplement)(J38);
rewritten as 'a supplement'.

- (by+NUM+divisi ons+:)+NPP(J39); rewritten as
  'by NPP'.
- (scientific+literature+use)+(:+a+survey)(J40);
  rewritten as 'literature use study'.
- (for+the+expansion)+(of+medical+information)
  (J42); rewritten as 'for documentation in a
  particular discipline'.
- (through+improvement)+(in+methods)(J42); rewritten
  as 'through methods'.
- (of+the+scope)+(of+medical+documentation)(J42);
  rewritten as 'of documentation in a particular
  discipline'.
- growth+(of+chemical+literature)(J43); rewritten
  as 'literature in a particular discipline'; this
  rewriting is concatenated with the code for (:+
  past+,+present+,+and+future)that follows, and is
  preserved as the final rewriting.
- support+(is+needed)(J43); rewritten as 'Verb,
  need+support'.
- (this+is)+(a+report)(J44); rewritten as 'a study'.
- (an+account)+(is+given)(J45); rewritten as
  'a study', the rewriting already assigned to
  'an account' in G.2.
- (greater+variety)+(in+collections)(J45); rewritten
  as 'collections'.
- (is+   +possible)+(nor+desirable)(J47);
  rewritten as 'verb, negation, possible+desirable';
  this string could have been more consistently
  processed as a conjunction, but this was not
  possible because the dictionary code which was
  assigned to 'nor' (XNOR) does not represent
  this word as a conjunction (C).
- (countless+efforts)+(have+been+made)(J49);
  rewritten as 'efforts'.
- (some+of+the+requirements+and+advantages)+
  (of+classification+systems)(J49); rewritten as
  'documentary languages'.
- (to+many+new+areas)+(such+as+high+capacity+
  storage+and+retrieval+and+data+management)(J50);
  rewritten as 'to documentation+related science',
  a rewriting already assigned to "(such as) high
  capacity storage and retrieval and data mana-
  gement" in G.2.

2.4.6. G.5

G.5 uses the same computer programmes as G.2. It concatenates in succession strings of six, five, four, three, and two codes and rewrites them as a single code. Like in the presentation of G.2, I shall just give a few examples of the rules of G.5, because a complete list would amount to quoting practically the whole of the experimental corpus. Only the strings of 6 codes will all be quoted below ; in total, 8 examples. Some of the examples not given in this section will be presented in later sections to explain the workings of other 'grammars'.

I shall present the examples by quoting first the object-language string, then the paraphrase of the corresponding codes, and lastly the paraphrase of the rewriting of the codes as a single code. Here again, I wish to stress the fact that the paraphrases of the codes are presented as problems generated by my experiment.

One general point which needs to be made about G.5 is the following. Whenever a conjunction or a comma stands in the way of a concatenation which could be performed in G.5, this concatenation is not performed and is left over to a later 'grammar' (G.7) operating after the necessary 'conjunction' reduction (in G.6). This is best explained by means of a few examples. Consider the following quotation from J12 : "binary-octal designations are established for theorems, systems of logic and for statements in propositional calculus". Before the operation of G.5, this string is concatenated as follows: (binary-octal+designations+are+established)(for+theorems+, +systems)+(of+logic)and(for+statements)+(in+propositional+ calculus). The "and" stops the concatenation in G.5,

which merges together the codes as shown by the following
bracketings : (Binary-octal+designations+are+established+for
+theorems+,+systems+of+logic)and(for+statements+in.proposi-
tional+ calculus). The string "for statements in proposi-
tional calculus" is conjoined with what precedes in G.6.
Similarly, in J32, the processing of a when-clause is delayed
until G7 because of the occurrence of a conjunction : "when
the problems of input and of the automatic generation of
classification information are solved". A comma also stops the
processing of "(the author) draws, from reports by NPP on..."
(J40) in G.5. With the qualification just made, the strings
processed in G.5 fall under the grammatical types listed below
in (1) to (1v).

I must also point out that some inconsistencies have
crept into th experiment. One of these may be mentioned here;
it concerns the treatment of subordinate and relative clauses.
Some of these are treated as illustrated under heading (111).
Others are treated as right adjuncts of nouns : for instance,
the 6-code example of J9 given below under heading (1). On
the basis of the experience gained during the experiment, I
am unable to say which of these two solutions is to be
preferred.

(1) Verbs, adjectives, cogiate nominals and nouns+strings
    following.
    The strings brought under this heading are partly equi-
    valent to Harris's "object of the verb" category (written
    as omega : Harris 1965, 30-33), and partly to his PN
    (preposition+noun) right adjuncts of N (op.cit., 37).
    Thus, for Harris "his description of the experiment" is the
    Object of "reviewed" and "of the experiment" the Object of
    "description" in "I reviewed his description of the
    experiment" (Harris, loc. cit.), and what is underlined in

Harris's example (i.e. everything except the subject) would
be concatenated in G.5. The right adjunct "in the painting"
in "the colors in the painting" would also be concatenated
with the preceding string "the colors" at this stage
(Harris, loc. cit.). The strings following verbs and
cognate nominals, however, may also consist of what Harris
would presumably analyze as subjects : for instance, a
by-phrase after a verb or a noun.

This is the place to mention a difficulty which I
have experienced with strings consisting of Noun+PN right
adjunct, which are processed   at this stage. The
difficulty can be explained in Stockwell et al.'s terms.
It has to do with a distinction made by Chomsky 1964
between an NP+Prep+N construction, "such as a house in
the woods", and "a reduced relative, such as that book
on the table" (Stockwell et al. 1968, 489). I have
attempted to incorporate this distinction into the order
of processing in the experiment. Thus, the concatenation
of "problem" or "problems" with "in communicating Russian
science" (J3) and with "of preparing an inverted file"
(J21) has been delayed until G.7, because parallel
expressions such as "problems confronted in..." (J21) and
"problems which would be encountered in..." (J29) are
processed in G.7. Compare the treatment of the string
"(the annotations) are written in the language used in
the pertinent geographical region :..." (J25) :
the concatenations (are+written+in+the+language)and(used
+in+the+pertinent+geographical+region+:...) are performed
in G.5 and the resulting codes. (written in the language)
and (of the region) are later concatenated in G.7.
Occasionally, however, I have failed to make the appro-
priate decisions in this respect, for lack of reliable

detailed knowledge about the lexical items and the
transformations involved. For instance, "alphabetical
indexes" has been concatenated in G.5 with the following
"of institutions, subjects, and personnel" (J25). This
may be a mistake, since paraphrases such as "indexes
giving/that give..." are possible. Moreover, the conca-
tenation in G.5 of "some problems and questions" with
"of patent documentation in industry" (J11) is not
consistent with the treatment of the strings with "problem"
just quoted. The same is true of the concatenations (three
representative problems)+(of a manufacturer of fine
chemicals)(J4) and "(should attend)+(to the problem)+
(of encyclopedic organization)+(of their knowledge)"(J16)
in G.5.

EXAMPLES

6 codes : - contains+(all+of+the+document+codes)+
(for+all+of+the+documents)+(in+which)
+(the+term)+appears (J9); contains+
(document codes)+(for documents)+(in
which)+(the term)+appears; 'contains
documents codes'.

- (making+it+possible)+(to+search)+(for+
all+of+the+information)+(in+the+file)+
(of+interest)+(to+the+searcher)(J19);
(making possible)+(to search)+(for the
inf ation)+(in the store)+(of interest)
+(to the searcher); 'making possible
to search'.

- representing+(all+of+the+subunits)+
(of+information)+(in+all+of+the+documents)
+(in+the+memory)+(of+the+machine)(J21);
representing+(the units)+(of information)
+(in the documents)+(in the store)+
(of the machine); 'for the stored
information units'.

- (will+make+possible)+(low-cost+memories)
(of+billions+of+words)+(with+access)+
(to+any+part)+(in+a+few+seconds)(J32);
(will make possible)+(stores+quantitative,
economic specifications)+(with possible
use)+(to a part)+(in time units);
'will make searching possible'.

- (a+description)+(of+Hungary+'s+network)+
(of+technical+libraries)+(with+NPP)+
as+(its+core)(J45);(a study)+(of national
organization)+(of documentary institu-
tions)+(with NPP)+as+(its core);
'a study of national documentation'.

5 codes : - Lists+(service+bureaus)+(in+Canada+and+
the+United+States)+(by+province+or+
state+,+and+by+city)+(within+province+
or+state)(J30); lists+institutions+
(in international area)+(by region);
'verb, lists international institutions'.

4 codes : - report+(on+a+study)+(of+scientific+
communications)+(for+NPP)(J2); report
+(on a study)+(of documentation)+(for
NPP); 'study of a study of documenta-
tion for NPP'.

- (this+reviews)+(the+initial+reasoning+
and+thinking)+(behind+the+design+and+
construction)+(of+a+project)(J2);
studies+(the thinking)+(behind the
development)+(of a project); 'studies
the thinking behind the development
of a project'.

3 codes : - (was+planning)+(to+change+over)+
(to+a+mechanized+storage+and+retrieval+
system)(J1); (planned)+(to change over)+
(to a mechanized documentation system);
'planned to use a mechanical system'.

- (NUM+representative+problems)+(of+a+
manufacturer)+(of+fine+chemicals)(J4);
problems+(of a scientist)+(of topic of
a science);'problems of a scientist'.

- (house+organs+and+trade+publications)
  +as+(information+sources)(J41);
  (special documents)+as+(information
  sources); 'special document sources'.
  Note the treatment of '<u>as</u>' at this
  late stage, because of its possible
  status as subordinate conjunction ,
  and because it could not safely be
  concatenated with what precedes until
  G.5.

2 codes : - (a+general+description)+(of+NPP)(J1);
  (a study)+(of NPP); 'a study of NPP'.

- similar+(to+that)(J1); similar+(to
  that); 'similar to that'.

- used+(by+NPP)(J1); used+(by NPP);
  'used by NPP'.

- available+(from+the+center)(J1);
  available+(from the center);
  'available from the center'.

- (the+charges)+(for+each)(J1);
  (the charges)+(for each); 'the
  charges for each'.

- (tracing+and+identifying)+(NPP
  references)(J4); finding+documents;
  'documenting special documents'.

- (the+dissemination)+(of+scientific+
  information)(J5); (the dissemination)+
  (of information); 'documentation'.

(11) <u>Preposition+Verb+'Object'</u>.

'Object' may include a relative clause (see 111), as

in the first example given below. 'Verb' includes cognate
nouns.

EXAMPLES

6 codes : - (from+reports)+(by+NPP)+(PON/HOW)+
  (scientific+personnel)+gather+
  information(J40); (from studies)+
  (by NPP)+(on how)+scientists+gather+
  information; 'from study of user'.

<u>5 codes</u> : - (to+develop)+(automatic+classifications)
+(of+chemical+concepts+and+interrela-
tions)+(in+the+disclosures)+(of+the+
patent+literature)(J19); (to develop)+
(mechanical classifications)+(of
information in a particular discipline)
+(in documents)+(of type of literature);
'to develop mechanically documentary
language in a particular discipline'.

<u>4 codes</u> : - (for+improving)+(the+dissemination)+
(of+scientific+information)+(PAMONG+
U.S.+scientists+and+engineers)(J5);
(for Verb,Qualitative)+(the dissemina-
tion)+(of information)+(among national
scientists); 'for national documentation'.

<u>3 codes</u> : - (to+affect)+(the+behavior)+(of+
scientific+institutions)(J2); (to affect)
+(the behavior)+(of institutions);
'to affect institutions'.

- (with+a+discussion)+(of+the+problems)+
(of+selection+,+form+,+distribution+,
+acceptability+,+and+translation+
rights)(J3);(with a study)+(of the
problems)+(of the development of trans-
lations);'with a study of the develop-
ment of translations'.

- (in+searching)+(for+information)+(in+
the+medical+and+chemical+literature)
(J4); (in searching)+(for information)+
(in the literature of a particular
discipline); 'in documenting in a
particular discipline'.

- (in+addition+to+presenting+a+general+
review)+(of+the+scope+of+the+Federal+
Government+'s+role)+(in+research+and+
development)(J5);(in addition to a study)
+(of governmental role)+(in science);
'in addition to a study of governmental
role in science'.

<u>2 codes</u> : - (to+determine)+ways(J2); (to determine)
+ways; 'to develop ways'.

- (in+communicating)+(Russian+science)
(J3); (in communicating)+(national
science); 'in documenting national
information'.

(111) <u>Subordinate and relative clauses not processed in</u>
<u>preceding 'grammars', and the pronoun the relative</u>
<u>clauses are adjoined to, if any.</u>

This includes SOME expressions of the type given in (11).
In the experiment, there have been some inconsistencies
and hesitations about these expressions, particularly
expressions of purpose. Compare, for instance, the examples
of J1 and J5 given below; in the example of J1 (<u>6 codes</u>)
an expression of purpose ("for the purpose of...") is
concatenated with what precedes ("...established");
in the example of J5 (<u>2 codes</u>), an expression of purpose
("for improving...") has not been concatenated with what
precedes ("which has been developed by NPP"). There have
been similar hesitations with <u>for-to</u> expressions, as in
"necessary <u>for</u> the machine <u>to</u> search the pertinent terms
in the machine's memory" (J9). In the experiment, I have
adopted the transformationally wrong segmentation
(necessary for the machine)   (to search...). I also
wish to point out that <u>that</u>-clauses after verbs like
"conclude" have not been concatenated at this stage with
the preceding verb. Lastly, note that what Harris analyzes
as <u>PN</u> adjuncts of center strings (e.g. '<u>at this time</u>' :
Harris 1965, 39), is processed at this stage : in fact,
the example given below from J1 ("at the time of this
report") is paraphrasable as a subordinate clause :
"(at the time)(when)this report was written".

EXAMPLES

6 codes : - (which+was+established)+(for+the+
purpose+of+contributing)+(to+the+
development)+(of+science+and+technology)
+(by+collecting+,+processing+,+storing+,+
and+retrieving)+(scientific+information)
(J1); (which was established)+(for
contributing)+(to the development)+(of
science)+(by documenting); 'which
documents'.

- The following example constitutes an
inconsistency in the experiment.
Other concatenations of relative clauses
with what precedes have been performed
in G.7 : (the+bureau+'s+well-established
+NPP+information+retrieval+program)+(,+
in+which)+(the+individual+lines+and+
paragraphs)+(are+the+retrieval+units)+
employed(J48); (the institution's
retrieval program)+(in which)+(the text
parts)+(are documentation units)+used;
'institutional retrieval program'.

5 codes : - (that+are+responsible)+(for+the+
dissemination)+(of+recorded+information)
+(in+order+to+achieve+a+net+improvement)
+(in+scientific+productivity)(J2);
(that are responsible)+(for the dissemi-
nation)+(of information)+(to verb,
qualitative)+(in science); 'which
document'.

- (so+that)+(advantages+of+the+new+
techniques)+(for+automatic+storage+
or+retrieval)+(of+information)+(may+be+
exploited)(J16); (so that)+(the
methods)+(for mechanized documentation)
+(of information)+(may be used); 'to
use mechanized documentation methods'.

- so+(the+information)+(is+divided)+
(into+as+many+sections+as)+(there+are+
different+codes)(J21); so+(the informa-
tion)+(is divided)+(into as many parts
as)+(codes); 'so that the information
quantitatively compared to codes'.

<u>4 codes</u> : - unless+(a+syste? 'o+procedure)+
(is+developed)+(.+preparing)+(the+
file)(J21); unles..(a method)+(is
developed)+(for+developing)+(the store);
'unless a method is developed for deve-
loping the store.'

<u>3 codes</u> : - (as+far+as+costs)+(per+information+unit)
+(are+concerned)(J11); (as far as costs)
+(per information+unit)+(are+concerned);
're costs'.

- although+(the+classification)+(is+
variable)(J19); although+(the classifi-
cation)+(is variable); 'although the
documentary language is variable'.

- (which+h e+believes+should+be+kept+in
+mind)+(by+those+who+manage)+(science+
libraries)(J40); (which must be known)+
(by those who manage)+(documentary
institutions); 'which documentary
institutions must know '.

- (that+will+determine)+(the+appropriate+
placement)+(in+a+catalog)(J12); (that
will determine)+(placement)+(in a cata-
log); 'that determines cataloging'.

- when+(the+problems)+(of+input)(J32);
when+(the problems)+(of input); 'when
the problems of input'.

<u>2 codes</u> : - (at+that+time)+(of+this+report)(J1);
(at the time)+(of this study); 'at
time of study'.

- (,+which+is+augmented)+(by+a+NPP)(J46);
(which is augmented)+(by NPP); 'which
NPP'.

- (which+has+been+developed)+(by+NPP)(J5);
(which has been developed)+(by NPP);
'which NPP develop'.

- (+substantial+proportion+of+
that+which+may+be+written)+(in+a+
natural+language)(J12);(what may
be written)+(in a natural language);
'natural language'.

- (;since)+(all+functional+units+,+
  and+not+merely+the+filing+categories+,
  +can+be+punched)(J26); since+(documen-
  tary language units can be punched);
  'since documentary language units can
  be punched'.

2.4.7. G.6

G.6 uses the same computer programmes as G.1 and G.3.
It processes codes corresponding to two types of strings :
conjoined strings, and strings separated by a punctuation
mark that are not conjoined, but are treated as such by the
programmes. Since an account of the type of strings processed
at this stage requires fairly long quotations, I shall only
give a few examples.

To present the examples, I shall represent as a single
concatenation what may in fact be the result of several
passes in the machine (due to the workings of the computer
programmes : see, sections 2.4.2. and 2.4.4.). Thus when
several passes are involved, as is typically the case in
enumerations, I shall only give the paraphrase of the **final**
code and not of the intermediary codes.

In (i) I present the examples of conjoined strings,
and in (ii) the examples of strings separated by a punctua-
tion mark. In this fashion, I shall also provide examples
of the workings of G.5 which were not given in the preceding
section. In fact, most of the codes on which G.6 operates
result from concatenations performed in G.5.

Note that the following type of string is treated as a
conjunction : "an outline of..., with a discussion of..."
(J3). For reasons which are unclear to me, there is indeed
a paraphrase relation between the above construction and the
superficial conjunction "an outline of... AND a discussion
of..." (cp. Jespersen on the relations between conjunctions
and prepositions : Jespersen 1965, 89-90). Also note the
following type of string, which is also treated as a

conjunction : "concerned primarily with..., most of the material has broad applications in..." (J13). This is paraphrasable as "Most of the material is concerned with... and has broad applications...". A similar treatment has been adopted for "...(thereby) making..."(J19,not exemplified below), and for "summary of...including description of..." (J31, not exemplified below).

A weak point of this 'grammar' is undoubtedly the great variety in the structures of the conjuncts it concatenates. In some rules, the conjuncts have much the same structure; for instance, "supported by NPP, carried out by NPP" (J2). But this is far from always the case. Next to the examples given in the preceding paragraph, consider the following example, which is repeated below : "(when the problems of input) and (of the automatic generation of classification information)" (J32). I am afraid that such clearly unnatural segmentations are the price which needs to be paid if one wishes to analyze language data with a fairly simple system such as that presented here. Perhaps the shortcomings under discussion could have been remedied within this system, by increasing the number of 'grammars' and by specializing their tasks. For instance, one could contemplate the setting up of special 'grammars' to deal with subordinate and relative clauses. However, it has not been possible to explore such possibilities within the time available for our experiment.
(1) Conjoined strings.

<div align="center">EXAMPLES</div>

- (supported+by+NPP)+,+(carried+out+by+NPP)(J2);
  (supported by NPP)+(developed by NPP); 'developed by NPP'.

- (an+outline+of+NPP+scientific+translation+
program)+(with +a+discussion+of+the+problems+
of+selection+,+form+,+distribution+,+acceptabi-
lity+,+and+translation+rights)(J3); (a study
of institutional translation program)+(with a
study of the development of translations);
'a study of institutional translation program'.

- (searching+the+botanical+drug+literature+for+
folk+remedies)+;+(interpreting+and+determining+
the content+of+foreign+prescriptions)+(;+and)
+(tracing+and+identifying+NPP+references)(J4);
(documenting in a particular discipline)+(docu-
menting translations)+(documenting special
documents); 'documenting special documents in
a particular discipline'.

- (in+addition+to+presenting+a+general+review+
of+the+scope+of+the+Federal+Government+'s+
role+in+research+and+development)+(,+and)+(a+
review+of+the+character+of+the+Federal+
Government+'s+scientific+information+activities)
(J5); (in addition to a study of governmental
scientific activity)+(a study of national
governmental documentation); 'a study of
national governmental documentation'.

- (is+divided+into+the+following+parts+:+the+
present+situation+in+the+United+States)+;+
(some+of+the+reasons+behind+the+current+
problems)+;+(the+principal+scientific+and+
technical+information+activities+of+the+
Federal+Government)(J16); (includes national
situation)+problems+(national governmental
documentation); 'includes national governmental
documentation'.

- (additional+information+concerns+availability+
of+microfilm+and+photocopying+services+or+
photocopying+service)+(only+,)(the+subject+
of+primary+interest+to+the+center)+(,+and)+
(coverage+provided+over+pertinent+Soviet+
periodical+literature)(J7);(studies photocopying
services)+(the discipline of the institution)+
(the documents); 'studies institution, disci-
pline, photocopying'.

- (a+general+discussion+of+the+present+state+
  of+biological+literature)+,+(with+emphasis+
  on+scientific+nomenclature)+,+(the+life+span+
  of+biological+literature)+(,+and)+(the+storage+
  and+retrieval+of+biological+specimens)(J10);
  (a study of the situation of literature in a
  particular discipline)+(with a study of the
  language of a particular discipline)+(the
  life of the literature in a particular disci-
  pline)+(documentation in a particular disci-
  pline); 'a study of the literature in a
  particular discipline'.

- (binary-octal+designations+are+established+for+
  theorems+,+systems+of+logic)and(for+statements
  +in+propositional+calculus)(J12); (develops
  codes for science related to documentation)
  and(for science related to documentation);
  'develops codes for science related to docu-
  mentation'.

- (concerned+   +with+chemical+information+
  storage+and+retrieval)+,+(most+of+the+material+
  has+broad+applications+in+other+disciplines)
  (J13); (studying documentation in a particular
  discipline)+(can be used in other disciplines);
  'studies documentation in a particular
  discipline'.

- (working+in+the+following+NUM+areas+:+coding+
  for+mechanical+searching+systems)+;+(retrieval+
  and+reproduction)+;+(equipment+for+information+
  storage)+;+(indexing+,+cataloging+,+and+
  classification)+;+(translation+by+mechanical+
  means)+;+(production+and+dissemination+of+infor-
  mation)+;+(use+of+information)+and+(user+
  requirements)+;+(theoretical+studies)(J14);
  (studying coding)+(retrieval and reproduction)+
  equipment+(documentation languages and cata-
  loging)+(mechanical translation)+documentation+
  (information use)+(user needs)+(theoretical
  studies); 'studying documentation, equipment
  and mechanical translation'.

- (based+on+the+results+of+this+study+,+recom-
mendations+were+made+for+application+of+
these+findings+to+existing+and+proposed+
systems)+(,+and)+(for+further+research+efforts+
in+the+evaluation+of+scientific+information+
retrieval+systems)(J17); (suggestions for
study of systems)[see(11)below]+(for study
of study of documentation systems); 'suggests
study of study of documentation systems'.

- (the+universal+code+of+science)+ and+(machine+
language)(J24); (the universal documentary
language)and(machine language); 'universal
documentary language and machine language'.

- (when+the+problems+of+input)+and+(of+the+
automatic+generation+of+classification+informa-
tion)(J32); (when problems of input)+(of
mechanical indexing); 'when problems of input
and of mechanical indexing'.

- (a+description+of+the+Automatic+Microfilm+
Information+System)+and+(how+it+works)(J36);
(study of the Microfilm System)+(its workings);
'study of the Microfilm System'.

- (description+of+the+Institute+,+founded in NUM)
+(,+and)+(its+services)(J39); (study of a
specific institution)[see(11)below]+(its
services); 'study of a specific institution'.

(11) Non-conjoined strings separated by a punctuation mark.

EXAMPLES

- (at+the+time+of+this+report)+,+(the+center)
(J1); (at time of study)+(the institution);
'the institution'.

- (binary-octal+designations+are+established+
for+theorems+,+systems+of+logic+and+for+
statements+in+propositional+calculus)+,
+(derived+from+certain+of+their+basic+
characteristics)(J12); (develops codes for
science related to documentation)[see(1)
above]+(from qualitative information);
'develop codes for science related to
documentation'.

- (presented+at+NPP+'s+NUM+NPP)+ ,+(held+in+
  March+NUM)(J15); (presented at NPP)+(held in a
  certain year); 'presented at NPP'.

- (based+on+the+results+of+this+study)+,+(recom-
  mendations+were+made+for+application+of+these+
  findings+to+existing+and+proposed+systems)(J17);
  (using the results of the study)+(suggestions
  for systems); 'suggestions for study of systems'.

- (description+of+the+Institute)+,+(founded in
  NUM)(J39); (study of a specific institution)+
  (developed in NUM); 'study of a specific
  institution'.

- draws+,+(from+reports+by+NPP+on+the +ways+
  in+which+scientifi:+personnel+gather+informa-
  tion)+,+(some+observations+as+to+certain+
  general+practices+and+ ctitudes+among+scientists)
  (J40); draws+(from user study)+(studies of
  scientists); 'studies user'.

- (if+U.S.+and+U.S.S.R.+rates+of+growth+continue+
  unchanged)+,+(the+U.S.S.R.)(J43); (if interna-
  tional action does not change)+(the specific
  country); 'the specific country'.

2.4.8. G.7

G.7 works with the same computer programmes as G.2 and G.5 and concatenates successively strings of six, five, four, three and two codes. Its main objective is to arrive at <u>a single code for each sentence</u>, whether articulate, semi-articulate, or inarticulate. When it fails to do so, this is due to the presence of a code for conjunction, and the ultimate reduction is performed in G.8.

Grammatically, the reduction or concatenation in each sentence of all the codes obtained in the preceding grammars to a single code involves various types of strings. These types are given below in (36), with a few examples. In (37), I give additional examples, presenting in succession strings of six, five, four, three, and two codes. No complete lists will be given for the strings of 5, 4, 3 and 2 codes. The only example of a string of six codes is given below. To present the examples of (37), I shall adopt the same conventions as those used in the preceding two sections. Most of the examples cannot be listed under the various grammatical types of strings given below, because many of the concatenations of G.7 involve more than one of these types. For instance, all the codes corresponding to the following bracketed strings are concatenated at this stage : (although this system was designed to be used with edge-punched, hand-sorted cards) (the author) (indicates that) (it can be modified for center-punched card) (which would permit machine sorting)(J26).

Here again, a more systematic (i.e. grammatically more motivated) approach is conceivable, but only at the cost of increasing the number of 'grammars' and specializing their function. ·

Lastly, note the rewriting as X, equivalent to 'null', when the information of a string is found unnecessary for the purpose of indexing : in the example of J5 given below, and in "(based on papers)+(presented by NPP at national meetings from NUM to NUM)"(J13).

(36)

GRAMMATICAL TYPES OF STRINGS CONCATENATED IN G.7

(i) <u>Subject+Predicate(in the traditional sense)</u>.

The only subjects not processed at this stage are tnose which, rightly or not, were processed in G.O (e.g. "<u>he</u>+feels",J12) and in G.4 (e.g. "<u>the address</u>+is given", J7). An example of a subject concatenated with the predicate at this stage is "(the charges for each)+(are listed)"(J1).

(ii) <u>Relative clauses not already processed+the code that precedes</u>.

This includes strings of the type "problems+(in+communicating)"(J3), under the assumption that "in communicating" is relatable to a relative clause such as "problems (<u>that were</u>) encountered in...". For instance : "(A+ general+description+of+NPP)+(which+was+established+for+ the+purpose+of+contributing+to+the+development+of+ science+and+technology+by+collecting+,+processing+,+ storing+,+and+retrieving+scientific+information)"(J1).

(111) <u>Strings equivalent to (11)</u>.

The equivalence is a matter of what can be referred to
in TG terms as "whiz" deletion (e.g. Langendoen 1970),
that is, the transformational deletion of relative
pronoun+<u>be</u>. For instance, "services+(available+from+
the+center)"(J1) is relatable to a construction with a
relative clause such as 'services <u>that are</u> available
from the center'. Similarly, constructions like "problem+
in..." : see the point made about these in connection with
G.5 (section 2.4.6.).

(1v) <u>Subordinate and 'that' clauses with what precedes</u>.
For instance : (concludes that)+(...)(J5), (...)+(so
that...)(J16).

(v) <u>Various other strings which were not concatenated before
+ what precedes or follows</u>.
First, "adjuncts of center strings", when no comma
follows ("in this file" in sentence-initial position
in J9) or when the comma has already been concatenated
with what precedes ("in+this+system+," in J19). Second,
strings mentioned under heading (11) in section 2.4.6.
(Preposition+Verb+'Object', concatenated in G.5) :
in particular, <u>by</u>-expressions ("<u>by</u> viewing it as...",J12),
and purpose ("<u>to</u> determine ways", J2) or relatable
expressions ("(procedure)<u>for</u>...",J12). Compare, for
instance, "...criteria and procedures <u>FOR</u> evaluating..."
and "criteria, and measures of performance <u>which might</u>
<u>be used TO</u> evaluate..." (J17). Third, strings separated
by a semi-colon that have not yet been concatenated : see
example of J29, below.

(57)

EXAMPLES

<u>6 codes</u> : - (in+NUM)+(the+NUM+articles)+abstracted
+came+(from+NUM+countries)+(in+NUM+
languages)(J43); (in NUM)+(the
documents)+abstracted+came+(from
countries)+(in languages); 'abstracting
of international documents'.

<u>5 codes</u> : - (this+reviews+the+initial+reasoning+
and+thinking+behind+the+design+and+
construction+of+a+project)+(supported
+by+NPP+,+carried+out+by+NPP)+(to+
determine+ways)+(to+affect+the+behavior+
of+scientific+institutions)+(that+are+
responsible+for+the+dissemination+of+
recorded+information+in+order+to+
achieve+a+net+improvement+in+scientific+
productivity)(J2); (studies the thinking
behind the development of a project)+
(developed by NPP)+(to develop ways)+
(to affect institutions)+(which document);
'studies thinking behind development
of a study of documentation institu-
tions'.

- (the+writer)+(feels+that)+(no+large+
scale+information+retrieval+system)+
(can+be+a+success)+(unless+it+can+handle
+searching+for+complex+relationships)
(J20); (the writer)+(feels that)+(no
documentation system)+(can+qualitative
information)+(unless it can search for
syntactic relations); 'need to search
for syntax'.

<u>4 codes</u> : - (at+the+time+of+this+report+,+the+
center)+(was+planning+to+change+over+
to+a+mechanized+storage+and+retrieval+
system)+(similar+to+that)+(used+by+NPP)
(J1); (the center)+(planned to use a
mechanical system)+(similar to that)
+(used by NPP); 'institution plans use
of mechanical system'.

- (the+author)+(concludes+that)+(the+
diseemination+of+scientific+informa-
tion)+(must+be+recognized+as+an+
integral+part+of+research+and+develop-
ment)(J5); (the author)+(says that)+
(documentation)+(must be viewed as
research); X.

3 codes : - (in+this+file)+(each+term+record)+
(contains+all+of+the+document+codes+
+for+all+of+the+documents+in+which+
the+term+appears)(J9); (in this
store)+(each term record)+(includes
document codes); 'term store includes
document codes'. Note that here the
relative clause had been - rightly or
wrongly - processed in G.5.

- (suggests+an+approach+to+the+cataloging+
of+the+entire+body+of+written+knowledge)
+(by+viewing+it+as+a+finite+series+of+
conjoined+propositions)+(which+may+be+
ordered+according+to+their+logical+
structure)(J12); (suggests theory of
indexing)+(by viewing it as topic of
science related to documentation)+
(which may be organized according to
science related to documentation);
'theory of indexing based on science
related to documentation'.

- (psychologists)+(should+attend+   +
to+the+problem+of+encyclopedic+organiza-
tion+of+their+knowledge+and+codification
+of+the+methods+,+measures+and+results
+of+psychology)+(so+that+advantages+of
+the+new+techniques+for+automatic+
storage+or+retrieval+of+information
+may+be+exploited)(J16); scientists+
(must study indexing in a particular
discipline)+(to use mechanizdd documen-
tation methods); 'indexing in a
particular discipline'.

2 codes : - problems+(in+communicating+Russian+
science)(J3); problems+(in documenting
national information); 'problems in
documenting'.

- (this+discussion+of+United+States+
scientific+and+technical+information+
services)+(is+divided+into+the+following
+parts+:+the+present+situation+in+the+
United+States+;+some+of+the+reasons+
behind+the+current+problems+;+the+
principal+scientific+and+technical+
information+activities+of+the+Federal+
Government)(J6); (study of national
documentation)+(includes national
governmental documentation); 'study
of national documentation'.

- (a+suggested+program)+(to+remedy+this+
country+'s+scientific+and+technical+
information+ills)(J6); (a suggested
program)+(to solve problems of national
documentation); 'suggests program for
national documentation'.

- (toward+a+procedure)+(for+logically+
cataloging+knowledge)(J12); (toward
a method)(for indexing by means of a
science related to documentation);
'theory of indexing by means of a
science related to documentation'.

- (lists+   +NUM+individuals+or+
organizations)+[rest of J14]; (lists
men and institutions)+(studying documen-
tation, equipment, and mechanical
translation)[see paraphrase of code
given in the preceding section, 2.4.▉];
'lists scientists studying documentation'.

- (the+computer+research+program+of+NPP)+
(is+described)(J18); (the computer
program of NPP)+(is studied); 'computer
program of an institution'.

- (classification+systems)+used (J25);
(documentary languages)+used; 'documen-
tary languages'.

- (systems+of+scientific+and+technical+
information+services)+(:+long+range+
planning)(J29); (systems of documenta-
tion services)+planning; 'future
development of documentation systems'.

- (the+simplest)+(being+a+small+,+
  manually+operated+,+portable+unit)
  (J35); (the qualitative+-est)(being
  manually operated machine); 'being
  manually operated machine'.

2.4.9. G.8

Like G.6, this last 'grammar' uses the same computer
programmes as G.1 and G.3, and it processes codes correspond-
ing to conjoined strings. The only difference with G.6 is that
a slight modification of the programmes makes it possible
to treat the codes for period (S) and for colon (KL) as
conjunctions, in the same way as are the codes for conjunc-
tions (such as 'and'), semi-colon, comma, and for the result
of concatenations performed in G.0 (e.g. semi-colon+'and').
Consequently, sentences in texts are treated as though they
were conjoined; the theoretical foundations of this approach
are examined in Part III.

This final stage of the processing raises the problem
of coding in all its magnitude. The codes adopted present
various deficiencies which are due to the lack of reliable
knowledge on discourse structure and to additional diffi-
culties inherent to the present experiment : the problem of
the loss of information in the indexing process, and the
need inherent to computation for the codes not to exceed
a limited number of characters specified in advance in the
computer programmes.

Contrary to what has been done in the presentation of
G.6, I shall give a paraphrase of the intermediary codes,
because this is necessary to present a crucial aspect of
the problem of discourse structure. As will be remembered,
the computer programmes used in G.8 work as follows : in a
conjunction AcBcC, A is first concatenated with B, then
A+B is concatenated with C. In G.8, A is usually the code for
a title, B the code for the first sentence of an abstract,
C the code for the second sentence, etc. The intermediary

codes - namely A+B, A+B+C, etc. - raise the problem of
discourse structure in terms of how to state the relations
between the title and the first sentence, between these and
the second sentence, etc.

About the final codes - that is, the codes that serve
to represent each abstract as a whole, I must point out
that some of them consist of two parts. One part, which is
capitalized in the paraphrases presented below, is a label
standing for one of the main headings of Gardin-Lévy (Appendix
2). The other part, which is written in small letters in the
paraphrases, consists of some additional information which
would ideally have to be expressed by means of more specific
descriptors. This, however, has not been attempted here,
since I only make use of the main headings of Gardin-Lévy
in the experiment. The additional information is therefore
just extracted from the texts, and juxtaposed to the Gardin-
Lévy headings used as descriptors. An example is the code
assigned to J1 in the experiment : CENTREDOCPLNMEC. It can
be paraphrased as 'DOCUMENTATION CENTRE planning mechaniza-
tion'. The capitalized part of this paraphrase corresponds
to Gardin-Lévy's heading 6. "scientific and technical
information institutions (studied individually)", while
"planning mechanization" is additional information drawn
from the abstract in question.

As far as discourse structure is concerned, I suggest
that the abstracts of the experimental corpus present various
types of situations which can be tabulated as in (37) below.
The first column indicates which parts of an abstract I
believe to be 'informative' or not, that is, to correspond or
not (partially or completely) to the topic or topics of the
abstract in question : respectively, the title (T),

sentence 1, 2, etc. (S1, S2, etc.). I shall try to explain more precisely what I mean by 'uninformative' below, by presenting a few examples. The 'uninformative' strings are given between brackets. In the second column I indicate which sentence or sentences are relatable to the 'informative' titles. In the third and fourth columns, I indicate the topics and corresponding strings for abstracts which have been assigned more than one descriptor. In the fifth column, I have marked the abstracts which are discussed, or used as examples of conjoined strings in the rest of this section.

(37)

Nr of Abstract \ Types	'Informative' parts:	Informative'title relatable to:	Abstract with more than 1 topic		Abstract discussed or exemplified below
1	T,S1,S2,S3	S1,S2,S3			X
2	(T),S1				
3	(T),S1				X
4	T,S1	S1	FUNCT:T	SOURCES:S1	X
5	T,S1(S2)	S1			X
6	T,S1	S1			
7	T,S1,S2	S1,S2			
8	T,S1,S2	S1	PROF:T,S1	S CI:S2	
9	(T),S1,S2				
10	T,S1	S1			
11	T,S1,S2,S3	S3	FUNCT:S3	EQUIP:S1,S2	X
12	T+whole abstract	whole abstr. except S3	CLASS: whole except S3	COD:S3	
13	T,(S1),(S2),S3	S3			X
14	T,S1	S1			
15	T,(S1),S2	S2			
16	T,(S1), S2,S3,S4	S2,S3,S4			
17	T,S1,S2	S1,S2			
18	T,S1,S2	S1,S2			
19	T,S1,S2,(S3), S4	S1,S2,S4			
20	T,S1,S2	S1,S2			
21	T+whole abstract	whole abstract			

Nr of Abstract / Types:	'Informative' parts:	'Informative' title relatable to:	Abstract with more than 1 topic	Abstract discussed or exemplified below
22	T +whole abstract	whole abstract		
23	(T),(S1),S2			X
24	T,S1,(S2),S3	S1, S3		
25	T,T,S1,(S2),S3	S1,S3		X
26	T+whole abstract	whole abstract		
27	T,S1,S2	S1,S2		
28	T,S1,S2	S1,S2		
29	T,S1,S2,(S3)	S1,S2		
30	T,S1	S1		
31	(T),S1			
32	(T)whole abstr.			X
33	T,S1,S2	S1,S2	SCI:T,S1, S2 CLASS:T, S1,S2	X
34	(T),(S1),S2			
35	T+whole abstr.	whole abstract	HAND EQUIP T+whole abstract	
36	T+whole abstr.	whole abstract	HAND:T+ whole abstr. EQUIP:T, S1	
37	T+whole abstr.	whole abstract		
38	T,(S1), remaining sentences	all sentences except S1		X
39	T,S1,S2	S1,S2		
40	T,S1	S1		
41	T,S1,S2,S3	S1,S2,S3	FUNCT:T S1 SOURCES: T,S1,S2, S3	
42	T,S1,S2	S1,S2		

Types: Nr of Abstract	'Informative' parts:	'Informative'title relatable to:	Abstract with more than 1 topic	Abstract discussed or exemplified below
43	T,S1,S2,S3, S4,(S5),S6	S2,S6	FUNCT:T,CENTER:S1 S2,S6 S3,S4	
44	T,S1,S2	S1,S2	FUNCT:T,CENTER:T, S1,S2 S1,S2	
45	(T),S1,S2			
46	T,S1,S2,S3 S4,S5	S2,S3,S4,S5	GEN:S1 GEN.ORG.: T,S2,S3, S4,S5	
47	T,S1,S2,S3, S4	S1,S2,S3,S4		
48	T,S1,S2	S1,S2		
49	T,S1,S2,S3	S1,S2,S3	SCI:T, HAND:T,S1 S1,S2,S3 S2,S3	
50	T,S1,S2, (S3)	S1,S2	SCO:S2 EQUIP:T, S1,S2	X

Before giving a few examples of concatenations of conjoined strings in abstracts, I wish to exemplify the above table by discussing briefly a few abstracts whose structure is noteworthy. One of them is J11,.In this abstract, the topic of the title ("<u>Some problems and questions of patent documentation in industry</u>": FUNCTIONING OF STI NETWORKS) is relatable to that of the LAST sentence : "<u>Five suggestions are advanced for a practical solution of patent documentation problems</u>". The other three sentences ("<u>...magnetic drums,...magnetic tape...information capacity</u>")have a topic (EQUIPMENT, storage devices) not explicitly related in the text to the other topic. J13 is an example of an abstract containing two 'uninformative' sentences, in the sense that they convey information irrelevant to indexing : "<u>This is a revised and enlarged edition of NPP. Based on papers presented by NPP at national meetings from NUM to NUM</u>". In J23, the 'uninformative' character of the title and of the first sentence is a consequence of the reduction of proper nouns to NPP : "<u>Basic directions of work at the Experimental Laboratory of Machine Translations</u>" (title). The topic (RELATED SCIENCES, machine translation) can, however, be inferred from the last sentence : "<u>One basic part of their program involves the use of an intermediate language between the input and output languages</u>". Let us also note the double title of J25. Both title and subtitle are relatable to the topic of the abstract, "GENERAL ORGANIZATION OF STI, national" : "<u>Archives, libraries, and documentation centers of Switzerland. Guide through Swiss documentation</u>". The last sentence of J29 is typically uninformative in that it is not relatable to any descriptor : "<u>The paper also includes discussion of problems which would be encountered in the development of such a system</u>". The problems arising in J32 have already been referred to. It is difficult to pin down all the reasons why this text, which is entitled

"The role of large memories in scientific communication" needs
to be represented by the descriptor GENERALITIES. One of them
is the use of the future tense as in "These memories will serve
as indexes to the deluge of technical literature", which makes
the abstract appear as a piece of science fiction rather than
as a truly scientific document. The whole of J33 has been
thought to be relatable to the topics INFORMATION SCIENCE
and CLASSIFICATION because the abstract concerns "a suggested
classification: for the literature of documentation" (title).
In J38, the topic (INFORMATION SCIENCE) is expressed in
general terms in the title ("Introduction to scientific docu-
mentation") and detailed in the abstract, except for the
'uninformative' first sentence : "The book comprises 4
chapters and a subject index". In J45, the title ("...
librarianship and documentation") has been labelled 'uninfor-
mative' because it is not clearly relatable to the abstract,
which is about "Hungary's network of technical libraries".
In J50, an 'uninformative' sentence of the type just quoted
from J38 occurs at the end :"It is classified and indexed".

In (β8), I present and discuss a few examples of strings
which are treated as conjunctions, mostly whole texts consist-
ing of a title and of one or more sentences. To present the
examples, I shall simplify the conventions adopted in the
preceding sections. In particular, I shall only use the + sign
(standing for concatenation) between natural language strings and
between corresponding codes that are processed in G.8.
In the examples given below, the natural language strings (with
only a few codes, such as NPP) are underlined, and the words
that are not underlined are paraphrases of the codes that get
concatenated in G.8.

(38)

- J 1 : (<u>NPP and its activities</u>)+(<u>A general description of NPP</u>
<u>which was established for the purpose of contributing</u>
<u>to the development of science...by collecting,processing,</u>
<u>storing,and retrieving scientific information</u>);
(the institution)+(the documentation institution);
'documentation centre'; (preceding string, coded as
'documentation centre')+(<u>At the time of this report,</u>
<u>the center was planning to change over to a mechanized</u>
<u>storage and retrieval system similar to that used by</u>
<u>NPP</u>); (documentation centre)+(institution plans use of
mechanical system); 'documentation centre planning
mechanization'; (preceding string, coded as 'documen-
tation centre planning mechanization')+(<u>services</u>
<u>available from the center</u>); (documentation centre
planning mechanization)+(activities of institution);
'documentation centre planning mechanization';
(preceding string, coded as 'center planning mechaniza-
tion')+(<u>the charges for each are listed</u>); (documentation
center planning mechanization)+(lists charges for
each); 'DOCUMENTATION CENTRE planning mechanization'.

COMMENTS. Note that the topic DOCUMENTATION CENTRE
is already assigned at the stage of the concatenation
of the code for the title with the code for the first
sentence. The second sentence adds the information
'planning mechanization'. The third and last sentence
("<u>services... and charges...</u>') gives information that
is included in the topic DOCUMENTATION CENTRE. Also
note that this last sentence has been wrongly segmented
:(<u>services available from the center</u>)and(<u>the charges</u>
<u>for each are listed</u>). The unfortunate consequence of
this segmentation, which was performed in order to be

consistent with the system, is the following : the
code for "<u>services available from the center</u>" is first
concatenated with the code for all that precedes in
the abstract ('documentation centre planning mechaniza-
tion'), and the code obtained by this concatenation
('documentation centre planning mechanization') has
then to be concatenated with 'the charges for each
are listed'. This could have been avoided by perform-
ing the concatenation of subject (<u>services available
from the center and the charges for each</u>) and predicate
(<u>are listed</u>) AFTER the operation of an additional
grammar for concatenating conjunctions of the type
"<u>services available from the center AND the charges
for each</u>".

- J 3 : (<u>Problems in communicating Russian science</u>)+(<u>An
outline of NPP scientific translation program, with
a discussion of the problems of selection, form,
distribution, acceptability, and translation rights</u>);
(Problems of documenting national information)+(A
study of an institution's translation program);
'TRANSLATION CENTRE'.

COMMENTS. In this abstract, as in J2, the meaning
of the title is vaguer than that of the abstract.
The topic can only be established by concatenating
the code for the title with the code for the abstract
which, in this case, consists of a single sentence.

- J 4 : (<u>Searching medicinal chemical literature</u>)+(<u>Three
representative problems of a manufacturer of fine
chemicals in searching for information in the medical
and chemical literature</u>); (documentation in a parti-
cular discipline)+(documenting the user in a parti-
cular discipline); 'functioning of STI networks,

in a particular discipline+user'; (preceding string,
coded as above)+(<u>searching the botanical drug
literature for folk remedies; interpreting and
determining the content of foreign prescriptions; and
tracing and identifying NPP references</u>); (functioning
of STI networks in a particular discipline+user)+
(documenting particular documents in a particular
discipline); 'FUNCTIONING OF STI NETWORKS,+user,
SOURCES in a particular discipline'.

COMMENTS. Here, the title already expresses the
first topic (FUNCT) and the first (inarticulate)
sentence ("<u>Three representative problems...</u>") conveys
additional information concerning a user ("<u>manufacturer</u>").
The enumeration following ("<u>searching the botanical
drug literature...</u>") exp esses the second topic
(SOURCES) because it refers to special types of
documents : "<u>botanical drug literature</u>","<u>foreign
prescriptions</u>","<u>NPP references</u>". Both the enumeration
and what precedes convey information concerning a
particular discipline ("<u>chemical</u>", etc.) and, in the
final code for the whole abstract this information
has been added to the headings of Gardin-Lévy, (cp.
'planning mechanization' in J1). Lastly, note that
the enumeration just mentioned is concatenated with
what precedes on the basis of the occurrence of a
colon; as already pointed out, colons are regarded as
a type of conjunction in G.8.

- J 5 : (<u>the Federal Government and U.S. scientific information</u>)
+(<u>in addition to presenting a general review of the
scope of the Federal Government's role in research and
development, and a review of the character of the
Federal Government's scientific information activities</u>);

(national governmental documentation)+(a study of
national governmental documentation); 'General
organization of STI+national+governmental'; (preceding
string,coded as above)+(this article presents a review
of the six-point program which has been developed by
NPP for improving the dissemination of scientific
information PAMONG U.S. scientists and engineers);
(General organisation of STI+national+governmental)+
(studies a program for national documentation);
'General organization of STI+national+governmental';
(preceding string, coded as above)+(the author
concludes that the dissemination of scientific infor-
mation must be recognized as an integral part of re-
search and development); (General organization of STI
+national+governmental)+X; 'GENERAL ORGANIZATION OF
STI+national+governmental'.

COMMENTS. The title of the abstract already
expresses the topic, which is also expressed in the
first sentence. In this sentence, it was not possible
to concatenate at an earlier stage the string
"this article presents..." with the long preceding
string "In addition to presenting a general review...".
The concatenation is performed at this stage on the
basis of the occurrence of a comma between the two
strings. While such a concatenation is not justifi-
able grammatically, it may be semantically.
Semantically, the comma between "In addition to
presenting..." and "the article presents..." acts
very much like a conjunction : 'presents...and
presents...'. The last sentence ("The author concludes
that...") was coded as X, that is, was regarded as
'uninformative' in the sense that the information

conveyed was found too trite to be worth indexing.
Another possibility would have been to represent it
by means of the descriptor GENERALITIES.

- J25 : (partial example): after the concatenation of title,
subtitle, and first sentence, the resulting code
(GENERAL ORGANIZATION OF STI, national) is concate-
nated with that for "<u>Descriptive data include</u>
<u>historical development, holdings, special collections</u>"
(Studies history and holdings); two further concate-
nations are necessary to process the rest of the
sentence :(classification systems used)+(user informa-
tion).

COMMENTS. This grammatically unacceptable concate-
nation of conjoined sentence parts with sentences
could only have been avoided by having an additional
'grammar', which would also have taken care of the
string "<u>services available from the center and the</u>
<u>charges for each</u>" in J1.

- J35 : (partial example) concatenation of the code for
"<u>When the document is to be reproduced</u>" and what
precedes (HANDLING+ EQUIPMENT) with the code for the
following sentence "<u>the appropriate card is selected...</u>"
on the basis of the occurrence of an intervening
comma. The resulting rewriting is again HANDLING+
EQUIPMENT.

COMMENTS. Such a treatment of a when-clause as
though it were conjoined with what precedes and what
follows is the only way I have found of taking care
of the intervening comma : "When...,...". Another
solution would have consisted of setting up a special
'grammar' for such subordinate clauses. Note, however,

that in this particular case, the sentence with a
when-clause is paraphrasable by means of a conjunction:
'the appropriate card is selected <u>and</u> the(a) document
is reproduced'. I shall not further discuss this point
here, as conjunction is one of the main topics dealt
with in Part III.

2.5. SAMPLE ANALYSES, AND CONCLUSIONS.

2.5.1 Complete sample analysis of one abstract.

In order to give a synthetic picture of the workings
of the system as explained in the preceding two chapters, I
now wish to present the complete analysis of one abstract -
namely J44 - as it was performed in the experiment and run
on the computer. I have selected J44 as an example for
various reasons. First, the coding used to analyze this
abstract is fairly readable and, although the text consists
of more than one sentence it is reasonably short. Besides,
it deals with two topics (GENERAL ORGANIZATION OF STI and
STI INSTITUTIONS) and one of these topics (GENERAL ORGANIZA-
TION) is by far the best represented in the experimental
corpus : 14 abstracts out of 50 have been represented by
means of the descriptor in question.

The presentation of the analysis of J44 will proceed
as follows. Let us note, first of all, that there are no
rules for this abstract in G.3 and in G. 6. Besides, to
simplify matters, I shall not subdivide the operations of
G.2, G.5 and G.7 into those involving 6, 5, 4, 3, and 2
codes, as is done in the computer programmes. Table (39) gives
a synthetic picture of the analysis of J44, omitting the
codes; the headings NPP and DIC stand respectively for proper-
noun reduction and dictionary-lookup. The codes are given
in tables (40) to (48). These tables serve to present the
dictionary-lookup operations, the reduction of proper nouns,
and the successive concatenations, from G.0 to G.8. Each
table consists of three columns. In the first table (40), which
exemplifies dictionary-lookup for proper nouns, the first

column contains the natural language strings, the second, the corresponding codes (in capitals)and the third, a paraphrase of these codes in small letters. In the second table (41), which exemplifies dictionary-lookup after the reduction of proper nouns to NPF, the first column gives the abstract (with a blank at the point where a reduced proper noun occurs), the second, the corresponding codes in capitals, and the third, a paraphrase of these codes in small letters. In the remaining tables (42 to 43) which show the workings of G.0, 0.1, G.2, G.4, G.5, G.7 and G.8, the first column gives the codes to be concatenated in capitals, the second (whenever this is necessary), a paraphrase of the codes in small letters, and the third, the code for the rewriting in capitals.

As will be noted in the illustrations given below, the codes can have up to 15 characters.

(39)
Analysis of J44 (codes omitted).

TEXT	NPP	DIC	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8
SOVIETx		x									
SCIENCE		x									
INFORMATION		x									
SERVICES		x									
.		x									
THIS		x									
IS		x									
A		x									
REPORT		x									
OF		x									
U.S.x		x									
ABSTRACTING		x									
AND		x									
INDEXING		x									
SPECIALISTS		x									
WHO		x									
OBSERVED		x									
THE		x									
OPERATIONS		x									
OF		x									
THE		x									
CENTRALIZED		x									
SCIENCE		x									
INFORMATION		x									
SERVICES		x									
OF		x									
THE											
ALL-UNIONx											
INSTITUTEx											
OF											
SCIENTIFICx											
AND											
TECHNICALx											
INFORMATIONx											
OF											
THE											
U.S.S.R.x											
.		x									
VINITIx	x	x									
IS		x									
PURPORTED		x									
TO		x									
BE		x									
THE		x									
x1ST		x									
COMPLETELY		x		Ø							
CENTRALIZED		x									
INFORMATION		x									
SERVICE		x									
IN		x									
THE											
WORLD		x									

ERIC

(40)

Dictionary lookup of proper nouns.

SOVIETx	NAM/GEO/NAT	Name,geographical,national
U.S.x	NAM/GEO/NAT	Name,geographical,national
THE ALL-UNIONx INSTITUTEx OF SCIENTIFICx AND TECHNICALx INFORMATIONx OF THE U.S.S.R.x	THE NOT CODED NAM/INS. POF NOT CODED C2 NOT CODED NAM/INF. POF THE NAM/GEO/NAT	Name, 'Institute' Preposition of Conjunction, and Name, 'Information'
VINITIx	NOT CODED	

(4)

Dictionary-lookup after the reduction of proper nouns.

SOVIETx	NAM/GEO/NAT	Name,geographical,national
SCIENCE	NSG/SCI.	Noun,singular,'science'
INFORMATION	NSG/INF.	Noun,singular,'information'
SERVICES	N2P/SER/VIC.	N plural or Verb,'service'
.	S	Period
THIS	XTHIS	'this'
IS	XIS	'is'
A	A	'a'
REPORT	N2S/STU/RPT.	N singular or Verb,study,'report'
OF	POF	Preposition of
U.S.x	NAM/GEO/NAT	Name,geographical,national
ABSTRACTING	VNG/DOC/ABS.	Ving,documentation,'abstract'
AND	C2	Conjunction,and
INDEXING	VNG/IND.	Ving,'index'
SPECIALISTS	NPL/DIS=AGT	N plural,agent of discipline
WHO	XWHO	'who'
OBSERVED	VED/STU/OBS.	Ved,study,'observe'
THE	THE	'the'
OPERATIONS	NPL/OPE/RTN.	N plural, 'operation'
OF	POF	Preposition of
THE	THE	'the'
CENTRALIZED	VED/ORG/NWK	Ved,institutional organization
SCIENCE	NSG/SCI.	N singular, 'science'
INFORMATION	NSG/INF.	N singular, 'information'
SERVICES	N2P/SER/VIC.	N plural or Verb, 'service'
OF	POF	Preposition of
	NPP	Proper noun
.	S	Period
	NPP	Proper noun
IS	XIS	'is'
PURPORTED	VED/SAY	Ved, say
TO	PTO	Preposition to
BE	XBE	'be'
THE	THE	'the'
x1ST	NUM	numerical information
COMPLETELY	ADV	Adverb
CENTRALIZED	VED/ORG/NWK	Ved,institutional organization
INFORMATION	NSG/INF.	Noun, singular, 'information'
SERVICE	N2S/SER/VIC.	N singular or Verb, 'service'
IN	PIN	Preposition in
THE	THE	'the'
WORLD	NSG/GEO/INT/VAG	Noun,geogr.,international,vague
.	S	Period

(⁴²)

G.0

PTO XBE	(see above)	XTOBE

(43)

G.1

VNG/DOC/ABS. C2 VNG/IND.	(see above)	VNG/DOC/ABS+IND

NAM/GEO/NAT NSG/SCI. NSG/INF. N2P/SER/VIC.	(see above)	NPL/NAT/DOC/NWK
XTHIS XIS		DXS/IS
A N2S/STU/RPT.		NSA/STU
POF NAM/GEO/NAT VNG/DOC/ABS+IND NPL/DIS=AGT	Preposition of Name,geographical,national Ving,document.,abstract and index N plural,agent of discipline	POFNPLNATDOCAGT
XWHO VED/STU/OBS.		WHO/VED/STU
THE NPL/OPE/RTN.		NPT/OPE/RTN
POF THE VED/ORG/NWK NSG/SCI. NSG/INF. N2P/SER/VIC.		POFTHENWKSERVIC
POF NPP		POF/NPP
XIS VED/SAY XTOBE		VED/VEW=PAS+SG
THE NUM VED/ORG/NWK NSG/INF. N2S/SER/VIC.		NSGNUMNWKSERVIC
PIN THE NSG/GEO/INT/VAG		PINNSTGEOINTVAG

(45)
G.4

| DXS/IS
NSA/STU | Deictic+is
N,singular,a,study | NSA/STU |

(46)
G.5

NSA/STU POFNPLNATDOCAGT	N,singular,a,study of national documentalists	NSASTUPOFDOCAGT
WHO/VED/STU NPT/OPE/RTN POFTHENWKSERVIC POF/NPP	who studied the operation of the organized service of Proper Noun	WHOSTUNPPNWKACT
VED/VEW=PAS+SG NSGNUMNWKSERVIC PINNSTGEOINTVAG	is regarded(viewed)as NUM organized service in the world	VEDVEWXASNWKSER

(47)
G.7

| NSASTUPOFDOCAGT
WHOSTUNPPNWKACT | a study of documentalists
who study Proper Noun organized
activities | NAM/INS/NWK/ACT |
| NPP
VEDVEWXASNWKSER | Proper Noun
is regarded as organized service | NAM/INS/NWK/ACT |

(48)
G.8

NPL/NAT/DOC/NWK S NAM/INS/NWK/ACT	N,plural,national documentation network Period Name,institution,organized activities	NETWRKNAT+CEN
NETWRKNAT+CEN S NAM/INS/NWK/ACT	national documentation network+ center	NETWRKNAT+CEN (final rewriting)
Comments: The final rewriting is a label corresponding to the headings "General organization of STI" and "STI institutions" of Gardin-Lévy (Appendix 2).		

For various reasons, I shall not present the complete computer printout for the analysis of J44 or of any other abstract. For one thing, the computer printout contains various repetitions which are of no linguistic interest. The computer was programmed to print over and over again the strings of codes to be processed. The advantages of doing this is that the workings of the programmes can be checked at any point in the processing. On the other hand, the resulting printouts for the analysis of each single text are extremely long and repetitive, and their reproduction would be a fairly expensive matter. Besides, the original computer printouts are filed at the Computing Centre of the University of Victoria, and the carbon copies which I have are not good enough for xeroxing.

2.5.2. Comments on the step by step concatenations.

The system presented above is based on step by step concatenations which eventually result in the reduction of a text to a single code. Here, I wish to say a few words of the pros and cons of such a system as I see them. The discussion will be fairly brief. A more elaborate discussion would have to be based on more experience than could be gained in the experiment reported here. In particular, it would have been necessary (but it was not possible for lack of time) to extend the experiment to a larger set of texts than the 50 abstracts of the experimental corpus, in order to see what changes would have been needed in the rules to permit such an extension.

I would regard the following as limitations or deficiencies of my system. First of all, I have not come across situations in which a conflict arises between two or more rules, and it is difficult to say how the system could handle such situations. Consider, for instance, the string "documentation centers". A rule of G.2 applied to this string in J25 has the effect of analyzing "centers" as a noun. Suppose we have another rule in G.2 which analyzes "centers" as a noun in the string "on documentation centers". The problem is how the system would analyze a sentence "This paper on documentation centers on problems of...", and how rules segmenting this sentence correctly could be reconciled with the others. If a conflict cannot be avoided between the rules, an obvious solution would be to set up one or more additional subroutines to re-process the codes and to replace them by other codes if the rules have resulted in incorrect segmentations. The existence of such an 'error correcting' subroutine in ACORN ("An Automated Coder of Report Narratives" in the field of medicine) is reported by Shapiro (1967, 157-8). As Shapiro

points out, conjunction is a frequent source of incorrect
segmentations or linkages in ACORN.

I am also unable to say how my experimental system would
handle lexical ambiguities of the type occurring in Kuno's
famous example :

(49)

TIME FLIES LIKE AN ARROW.
1. Time passes as quickly as an arrow.
2. A species of flies called "time flies" are
   fond of an arrow.
3. You shall time the flies which are like an
   arrow. (Or You shall, as quickly as an arrow,
   time the flies.) (Kuno 1966, 88).

I have not found any comparable cases of ambiguity in the
Janaske corpus.

Thirdly, the system does not display syntactic ambiguities
and imposes one interpretation on ambiguous strings : the
example given above in section 2.4.2 was "interpreting and
determining the content of foreign prescriptions" (J4).

Fourthly, the system has no transformational component
that would display explicitly relations between sentence types
or sentence constituents - either in the fashion proposed by
Harris and his disciples, or in the fashion proposed by the
Chomskyan tradition.

Lastly, a grammarian who has examined my examples closely
and noted the various inconsistencies which I have pointed out
or which appear from the examples will undoubtedly realize
that I have taken various liberties with string theory. More
generally, a grammarian may feel that it is often difficult
to trace direct connections between my experimental system
of concatenation rules and rules of any recognized linguistic
theory. I realize that my experiment is partly open to this

criticism. But I cannot accept it as a serious objection inso-
far as I feel that many of my rules which are not well
motivated linguistically involve problems that have been
left undiscussed or unsolved by current linguistic theories.
In particular, if one objects to the kind of 'text trees'
presented in the preceding section, I can only answer that I
am not familiar with any other attempt to set up text trees
and to mechanize the corresponding analyses for indexing
purposes.

Among what I would regard as advantages of my system, I
would point out the following. As a computational experiment,
the system has worked well from beginning to end. It has been
tested with a fairly large corpus. This corpus is certainly
not comparable in size to the 1016   abstracts used in the
Bély et al. experiment (Bély et al. 1970, 137), or perhaps
to the corpora processed in the SMART system (although I have
not seen figures published by Salton or co-workers concerning
the number of texts they have analyzed). But my corpus of 50
abstracts is a sizable one if compared to the mostly sentence-
bound experiments referred to by Simmons in the state-of-the-
art review quoted above.

It is also important to realize the 'modular' aspect
of my system. It consists of several parts which could be re-
arranged in a different order and can thus serve as a tool
for experimenting with problems of rule-ordering. Moreover,
other computer programmes can be added to the system, should
this turn out to be necessary. In particular, it would be
possible to improve on the system by providing for right-to-
left scans (such scans had been programmed for the experiments
reported in my M.A. thesis : J. Noël 1968, 172-3), and for
error-correcting procedures of the type mentioned in
Shapiro 1967.

I also wish to stress the importance of the multiple-
pass recognition system adopted in G.2, G.5 and G.7, by which
strings of 7,6,5,4,3, and 2 codes are processed in succession.

Lastly, the existence of 'grammars' dealing with con-
junctions after each of the other 'grammars' (except G.4)
is based on the assumption that the problem of conjunction
and the related problem of certain punctuation marks (such
as comma) constitute some of the most difficult problems
which mechanical analysis has to face. Various experts in
the field seem to agree with such an assumption : Sager
1967 (172), Coyaud (and Siot-Decauville 1967, 39, referring
to work on a Russian corpus in geometry by Leontieva), and
Shapiro 1967. In fact, I suggest to regard my 'grammars'
dealing with conjunctions as basic, and to regard the other
'grammars' as attempts to state conjoinability conditions.

2.5.3. Comments on the coding and on related problems of 'pragmatic competence'.

Throughout the presentation of my experiment, I have insisted on the problematic aspect of the coding, and on the difficulties I experienced in setting up the codes. These difficulties have practical, theoretical, and metatheoretical aspects.

The practical difficulties arise from two decisions which were made in the course of the experiment. One was to limit the number of characters available for the codes of any 'grammar' to a maximum of 15. Such a format provides for many possibilities if arbitrary codes are used : for instance, all the letters of the alphabet in all of the 15 positions available would provide for the coding of 26^{15} different situations. I found, however, that for the purposes of an experimental system it was necessary to use codes in clear, or at least codes consisting of mnemonic abbreviations of plain language words. But this decision had the effect of limiting severely the possibilities of the 15-character coding format. Particularly in the last 'grammars', some of the 15-character mnemonic codes which were used had the effect of imposing undue losses of infcrmation on the coding, or the mnemonic function of the codes had to be partly given up in order to avoid such losses of information. This is the reason why I have mostly presented code paraphrases rather than the actual codes in chapter 2.4.

The decision to use mnemonic codes is, I believe, fully
justified by theoretical and metatheoretical considerations.
First of all, this decision gives recognition to the fact
that little or no reliable and detailed knowledge is available
about lexical items in general, and about those occurring
in a corpus of documentation English like Janaske in particular.
I believe that the extent of our ignorance is such that it
cannot be remedied by a single author or even by teams of
linguists working for many years on a long-term project.
A computer count has revealed that the Janaske corpus
contains 6766 different word tokens (including capitalized
words) and about a thousand of these occur in the first
fifty abstracts which served as the experimental corpus.
In Part III, I have only been able to examine a small
portion of the words occurring in Janaske. This is only
natural considering that, for instance, a linguist like
Fillmore finds it necessary to devote a whole paper (actually,
two papers on the same topic : Fillmore 1969a and 1969b) to a
dozen of 'verbs of judging'.

The difficulties which I have experienced with the coding
have an additional, metatheoretical dimension if one considers
a few examples in the light of the concept of pragmatic
competence presented in Part I, or of Gardin's similar
concept of "double competence" :

> l'analyse sémantique d'un texte scientifique, fût-il
> déjà résumé, est une opération éminemment intelli-
> gente, qui exige une double compétence, sur le plan
> de la langue tout d'abord, mais aussi sur le plan
> de la pensée scientifique elle-même, puisqu'enfin
> l'on attend plus aujourd'hui d'un documentaliste
> omniscient qu'il soit capable de dégager indifférem-
> ment le sens d'un article de physique théorique
> ou de sociologie. La machine doit être instruite
> de la même manière dans ces deux ordres de compé-
> tence (J.G. Gardin's preface to Bély et al. 1970,
> xiv).

As examples of the problems I have in mind, I shall examine
a few passages from J12, J35 and J11 (see Appendix 3). For
convenience, these abstracts are again quoted below:

60)

Toward a procedure for logically
cataloging knowledge.
Suggests an approach to the cataloging of the entire
body of written knowledge by viewing it as a finite
series of conjoined propositions which may be ordered
according to their logical structure. An application
of this method to cataloging theorems of modern
logic is selected as particularly appropriate for il-
lustrative purposes. Binary-octal designations are
established for theorems, systems of logic and for
statements in propositional calculus, derived from
certain of their basic characteristics. The author
recognizes that not all statements can be expressed
in as neat a fashion as his examples from the field
of logic. He feels, however, that at least sub-
stantial proportion of that which may be written in
a natural language can be reduced to a standard
form that will determine the appropriate placement
in a catalog. (J12)

61)

Automatic selection and reproduction
of industrial documents.
The system described employs three established
techniques : microfilming, punched-cards, and
xerography. Documents to be held on file are
microfilmed. Each individual picture is cut off the
reel of microfilm and mounted over a special aper-
ture cut in a standard eighty-column punched-card.
Data identifying the microfilm are punched in
columns available on the card. When the document
is to be reproduced, the appropriate card is
selected either manually or mechanically from the
file of cards and placed in a special projector
mounted on a xerographic printer. The required
number of full-size documents is then printed auto-
matically at a speed of twenty feet a minute. The
system is particularly suited to the filing and
reproduction of engineering drawings--600,000
drawings can be recorded and filed in a space of
fifty square feet. Machines of varying complexity
are available for mounting microfilm on the cards,
the simplest being a small, manually operated,

portable unit. Details are given of the Copyflo high-
speed printer, and the way in which the system is
operated by the Westinghouse Electric Corporation
is described, with figures of savings achieved in
storage space and in costs of drawing-office supplies
and of reproduction drawings. (J35)

(52)

Some problems and questions of patent
documentation in industry.
A comparison is made among magnetic drums,
magnetic tape, magnetic wire, ferromagnetic
cores, punched-cards, punched tapes, and photo-
graphic tapes with respect to their access,
durability, storage and retrieval time, capacity,
and efficiency. It is concluded that the most
economic, as far as costs per information unit are
concerned, is magnetic tape. It also possesses the
greatest information capacity. Five suggestions are
advanced for a practical solution of patent docu-
mentation problems. (J11)

To begin with, consider the problem of coding as has been
done in the experiment, that is, as a descriptive problem.
Let us first examine the passages from J12 which are listed
below in (53), and in particular the verbs or verbal expres-
sions underlined in these passages :

(53)

- logically cataloging knowledge
- the cataloging of the entire body of written knowledge
  by viewing it as a finite series of conjoined propo-
  sitions which may be ordered according to their logical
  structure.
- not all statements can be expressed in as neat as
  fashion as his examples from the field of logic
- at least substantial proportion of that which may be
  written in a natural language can be reduced to a
  standard form that will determine the appropriate
  placement in a catalog.

It seems reasonable to regard these passages as semantically
equivalent, to capture this equivalence by assigning the
descriptor CLASSIFICATION AND INDEXING to J12 as has been
done in the experiment, and to regard the underlined expres-
sions as equivalent on the basis of this equivalence. But
even as a descriptive problem, I have found that the coding
of such equivalences could only be handled empirically and
left open to further research. One reason is that no linguistic
theory known to me would be able to display the semantic
equivalence (for instance, in terms of implication rules as
proposed by Leech 1969) that underlies the grammatical
diversity of the expressions under consideration.

Another reason is that the equivalence in question may
be a matter of pragmatic competence. Only an expert in
information science may be able to perceive the equivalence, and
it may hold only in the English of information science :
for instance, everyday English would equate "expressed in a
neat fashion" with "well written", etc., and not with
"cataloging", and only in the logician's or the linguist's
jargon could the expression in question be equated with a
verb like "formalize".

The problem of pragmatic competence arises most clearly
at the discourse level in J35, the first sentence of which
refers to a system based on techniques of "microfilming,
punched-cards, and xerography". From the following sentences,
it appears that one of the main characteristics of the system
in question is an extremely specific chronological sequence
in the application of the three techniques mentioned in the
first sentence. Here then, discourse structure appears to be
determined by the very structure of the "system" described, as
set up by the system's designers. Here again, my coding

could only be empiric, since I do not have these experts'
expertise.

To anticipate a discussion presented in Part III, let us
now assume that "for all conjunction" (including "and", and
some commas in written language) "a common topic is necessary"
(Robin Lakoff 1970, 55). If this assumption and that made
in the preceding paragraph are correct, then the semantic
equivalance or common topic in "microfilming, punched-cards,
and xerography" (J35) is a matter of pragmatic competence,
and the coding which needs to capture this common topic will
have to be based on the non-linguist's expertise just referred
to. This is also particularly clear in the following passage
f om J11 : "A comparison is made among magnetic drums,
magnetic tape, magnetic wire, ferromagnetic cores, punched-
cards, punched tapes, and photographic tapes with respect to
their access, durability, storage and retrieval time, capacity,
and efficiency". In order to capture the "common topic" in
the first enumeration, I have assigned the same code -
paraphrasable as "storage equipment" - to all the conjuncts
of this enumeration. My knowledge of the subject, however,
did not enable me to find a satisfactory label for the common
topic in the second enumeration ("with respect to..."):
the only sensible code I could arrive at (in G.3) is para-
phrasable as "with respect to some qualities". Moreover,
I would stress the following point. The type of expertise
involved here is one which concerns creative aspects of language
use : whereas a word like, say, (human) limb is equatable to
"legs and arms". it seems to me that a list of 'storage
devices' and of their properties in computer science is
open-ended, because a computer scientist can at any time
invent and name a new storage device, or discuss some new
property of such devices. In Part III, I shall not deal syste-
matically with problems of "symmetric" or reversible conjunctions
as illustrated above in J11, in order to concentrate on "asym-
metric" conjunction (R. Lakoff 1970).

2.5.4. Conclusions.

The mechanized indexing system under discussion was set up with a limited objective in mind. The purpose was to show that it is possible to mechanize the reduction of abstracts in a particular field to a single code standing for a descriptor in this field, and to do this by means of step by step concatenation rules based on an adaptation of string theory and its extension to discourse. This means that the system has been designed as a system for reducing surface structures (corresponding to abstracts) to surface structures (corresponding to descriptors). I am not acquainted with any similar attempt to set up purposefully a mechanized 'surface-to-surface' system. For this reason, it is difficult to compare my system with any other system for automatic indexing of abstracts. The Bély et al. experiment, in particular, which is the only one known to me that has underlying metatheoretical assumptions similar to mine, makes use of what might be referred to as deep structure representations based on the SYNTOL relations.

To conclude, I submit that the interest of my experiment lies in the gap between its full success as an experiment in computation, and its shortcomings with respect to the ideal metatheoretical requirements put forward in Part I : in particular, the ideal of a reversible procedure based on a linguistic study of pragmatic (or double) competence.

One theoretical deficiency of the system lies undoubtedly in the lack of elaboration of the rules, and in various inconsistencies discussed in the preceding chapter or appearing from the illustrations. One of Gardin's recent comments is applicable here : "les macro-mécanismes de l'analyse documentaire (...) sont des sortes de "court-circuits" d'une analyse plus fine du discours, au sens où l'entendent les

transformationalistes (Z.S. Harris et ses épigones aux
Etats-Unis, Maurice Gross en France, etc.)" (preface to Bély
et al. 1970, xvii).

The most serious deficiency, however, has to do with
problems referred to in the preceding section in connection
with the codes used in my experiment at the dictionary-lookup
stage and in the various successive 'grammars'. In the preced-
ing section, I have mentioned my difficulties with coding
conceived as a descriptive problem. But description is only
the crudest stage of theoretical investigation. To be
adequate, a mechanized indexing system needs to have predictive
power. The system under discussion, however crude it may be,
suggests the magnitude of the problem in the following terms :
to lend a mechanized indexing system predictive power one
would have to be able to enumerate, just for the recognition
of surface structures in texts, all the possible rules in at
least 9 'grammars' operating in succession. Moreover, if
what has been said in the preceding section is correct, the
setting up of such a generative system would have to rely
on what I have referred to as 'pragmatic' or "double" competence.

One might be tempted to consider that the need to predict
and state a great number of detailed rules in the various
'grammars' imposes undue rigidity on the system I propose.
I feel, on the contrary, that this feature of my system raises
the real problem of linguistic theory and of automatic indexing,
namely that of stating a finite number of rules that will
account for what Chomsky calls the creative aspect of language
(Chomsky 1965, 6). I also believe that, from the point of
view of linguistic theory, it does not make sense to have a
machine establish what a theory cannot predict, or to have
the machine 'discover' rules which the system designer has
failed to state.

The following, and last part of this thesis (Part III)
is based on the assumption that at least some of the short-
comings of my mechanized indexing experiment can be remedied
on the basis of a particular conception of semantic deep
structure. I assume that this type of semantic investigation
is a necessary addition to the theoretical study of the string
aspects of language, that such an addition is necessary to
serve the predictive purposes of a linguistic study of prag-
matic competence, and that a transformational apparatus can
eventually be set up to relate the semantic representations
to the surface representations. In this sense then, Part III
as a whole can be regarded as a conclusion of Part II.

FRANÇOISE & JACQUES NOËL-DAGNEAUX

Noël, Jacques, <u>A Semantic Analysis of Abstracts.
Around an experiment in mechanized indexing.</u>

University of Liège (Belgium), doctoral
dissertation, 1971-1972. Volume 2 (Part 5,
Appendices, Bibliography)

ALLÉE DE LA CENSE ROUGE, 50
(B-) 4250 OUGRÉE (SART-TILMAN)
BELGIQUE
Tél. : (04) 65.64.79

PROF. DR. J. NOËL, M.A.
VRIJE UNIVERSITEIT BRUSSEL
SECTIE GERMAANSE FILOLOGIE

262

PART III : PROBLEMS OF SEMANTIC REPRESENTATION

3.1. THEORETICAL BACKGROUND : CASE THEORY AND BEYOND.

3.1.1. Introduction and outline.

Although I have intended this part of my thesis to be a study of some aspects of English from a TG point of view, there are obvious differences between my preoccupations and those of the TG studies with which I am acquainted. Superficially, the difference can be characterized as follows. The TG linguists' examples consist by and large of sentences they invent : they range, let us say, from almost moronic three-word sentences about "John" or "Mary" to huge sentences by means of which the TG linguist attempts to throw some light on the grammatical or semantic possibilities of language. My examples, as far as possible, are drawn from a corpus of non-everyday English, and my ultimate preoccupation is with texts or discourses, rather than 'sentences' as the observable objects with which TG linguists are usually concer ed. Furthermore, the TG linguists' selection of problems is entirely determined by the state of TG research. A topic or problem is judged as interesting or, on the contrary, as trivial on just this basis. My selection of problems is only partly made on such a basis, and many of my problems are not yet in what current TG linguists would describe as a 'researchable' form. What I do share with TG linguists - and no doubt with many other students of language - is a conception of language as something systematic, the system of which is partly independent of the examples chosen for study. I believe that there would be no point in studying documentation English (nor in fact any language variety , including everyday language) unless one

can throw light on some linguistic universals, that is, on some points which can tell us what languages in general are like. I assume that my study of English abstracts in the field of documentation can serve such a purpose. The examples I have selected for this part of my study are drawn from the whole of the Janaske corpus, because I felt that my purpose would be better served in this way than by limiting myself to the experimental corpus of 50 abstracts dealt with in Part II.

More fundamentally, my preoccupations in these chapters on semantic representation can be summed up as follows :

(i) I have tried to strike a balance between two approaches to problem selection : the TG linguists' practice of inventing their examples (includi unacceptable sentences) on the sole basis of the state of TG theories, and the pre-Chomskyan practice of basing one's investigation on a selection of sentences which non-linguists would use or have actually used.

(ii) My ultimate concern is with extending the observational and theoretical tools set up by TG linguists (particularly by and around Fillmore) to the study of <u>texts</u>.

(iii) My main preoccupation is how the linguist can arrive at a semantic representation by comparing and manipulating a selection of observable data or 'words on the page'. In this, I follow a practice which has become more and more common among many TG linguists in recent years as a result of "an increasingly pessimistic attitude toward formalism" (Lee 1971, 11). Rather than attempt to write rules whose objective are beyond the powers of current theories, I have concentrated on the question of the nature of the linguist's <u>arguments</u> , because I believe this question to be a fundamental one. The tree diagrams

which I present should be regarded as a convenient (and,
I think, revealing) way of displaying the consequences of
I propose, rather than as statements of rules.

(iv) As my title suggests, I am concerned with semantic
representation as such. This means that I only deal with
what belongs the base of a generative linguistic theory
and that I do not attempt to set up transformations. My
claim is that the nature of the major problems I am
concerned with does not yet lend itself to a
transformational treatment. One of the consequences is that
the way in which this part of my thesis is organized does
not take account of current hypotheses about a transforma-
tional cycle (see, for instance Mc Cawley 1970).

(v) I have chosen to work within a theoretical framework set
up by Charles Fillmore and to attempt an extension of this
framework, chiefly on the basis of some recent work on
conjunction by Zellig Harris (1968) and Robin Lakoff
(1970). I believe, however, that some parts of my investiga-
tion could serve as a starting point for a comparative
study of what Fillmore's proposals and completing
or diverging frameworks (for instance, generative semantics)
could achieve when dealing with the same problems. In
particular, I go out from Fillmore's conception of a
methodological separation between phrase-structure rules
and lexical entries in the base of a generative theory of
language, a hypothesis inherited from Chomsky 1965.
To me, one of the primary tasks of case theory - as well as
one of its unique features - is that of correlating gram-
matical (and observable) facts such as what sentence
constituents can be chosen as subject, as object or as
prepositional phrase, with semantic facts concerning the
nature of these constituents. Indeed, this bringing together
of grammatical and semantic predictions constitutes one

of the major concerns of this part of my thesis - one
which parallels my concentration on linguistic argumenta-
tion (11.1).

(vi) My selection of problems is determined by a combination
of three considerations : the theoretical framework which
I have chosen; the nature of my data (a corpus of abstracts
and a corresponding metalanguage); and what I believe to
be the major shortcomings of the experiment presented in
Part II. Though I have not been able to present a detailed
semantic analysis of any abstract of the Janaske corpus or
of any entry of the Gardin-Levy concordance, I believe that
the problems I have selected are relevant to a description
of both the metalanguage and object-language under investiga-
tion, in accordance with the metatheoretical assumptions
of Part I.

(vii) I realize that all the problems I have selected for
discussion could have been treated more thoroughly, and
must admit that some important problems have not been
dealt with at all. For instance,I have no original proposal
to make about the question of nominalization', which is
an obvious problem in a .corpus of abstracts. My concentra-
tion on a few "cases" leaves little room for discussion of
other "cases" which a thorough semantic analysis of the
conceptual field of documentation would require. For one
thing, this thesis is long enough as it is. Moreover, I
have chosen to take the responsibility of dealing less
thoroughly than could have been done with a limited range
of problems, at the expense of others, rather than concen-
trate on a single point and ignore various others which I
believe to be related, even though current theories fail
to capture the relation.

(viii) Since my intention is to go beyond Fillmore's explora-
tion of "the elementary structure of the "propositional"
core of simple sentences" (Fillmore 1968b, 28-29),

it was to be expected that the linguistic literature
available to me proved, on the whole, to be of little
direct help for my descriptive purposes; the only notable
exceptions are some fundamental points made by   authors
mentioned in this section. For this reason, and also because
there is no point in duplicating various reviews and
discussions of TG linguistics mentioned in my bibliography
(including a paper of mine : J. Noël 1971), I do not
attempt to explain various well-known features or techni-
calities of TG linguistic theories in the present chapter
(3.1.). Nor do I attempt to give a complete account of
Fillmore's case theory and related frameworks; a critical
review of the literature on case would in itself require
a monograph (see the long bibliography given in Fillmore
1970c). The few technicalities about TG theory that are
directly relevant to my problems will be presented in due
time in the course of the discussion.

My discussion of problems of semantic representation falls
into five main chapters. The present chapter (3.1.) is entitled:
"Theoretical background : case theory and beyond". It consists
of three sections, in addition to the present introduction
(3.1.1.). The second section (3.1.2.) presents a brief account
of the phrase-structure, or constituent-structure subcomponent
of the base as set up by and around Fillmore. The third
section (3.1.3.) summarizes some of the most recent proposals
made by Fillmore about lexical entries. The last section
is intended to trace connections between two types of proposals
- a task which, to my knowledge, has so far not been attempted :
on the one hand, a hypothesis most thoroughly explored by
Annear Thompson according to which a sentence with a restrictive
or non-restrictive relative clause should be derived from a
pair of conjoined sentences; on the other hand, recent work
by Zellig Harris (1968) and Robin Lakoff (1970) on sentence

conjunction and discourses, as well as a few other relatable proposals. The discussion is centred on problems of identity : the partial identity which, according to Zellig Harris and Robin Lakoff, is required between conjoined sentences, the identity between noun phrases which needs to be stipulated in all known accounts of relative clause formation, and the broader problem of discourse cohesion. The major interest of Robin Lakoff's treatment of conjunction, as far as the purposes of this study are concerned, is the distinction she explores between <u>symmetric</u> or reversible sentence conjunction (as in 'John is a teacher and Mary is a chemist'), and <u>asymmetric</u> conjunction (as in '<u>Veni, vidi, vici</u>').

The second chapter (3.2.), which is entitled '<u>A discussion of...</u>' is the one which comes closest to the usual preoccupations of Fillmore and his followers. It deals with problems concerning expressions such as '<u>A discussion of -</u>', '<u>the author discusses-</u>', which are used in many abstracts to introduce a description of what the abstracted document is about. The four areas of concentration in this chapter are :

(i) arguments in favour of a locative analysis of, say, "this paper" in sentences of the type '<u>In this paper, the author discusses__</u>','<u>This paper discusses—</u>', etc.

(ii) arguments in favour of an analysis of x in sentences like 'The author discusses x' as the realization of a case which I propose to label the Topic case ,

(iii) an attempt to account for the semantic relatedness between sentences such as '<u>This paper discusses—</u>', '<u>In this paper, the author discusses—</u>', '<u>A brief discussion of—</u>', '<u>Summarizes—</u>', '<u>Included in this paper is a discussion of—</u>', '<u>This article is on——</u>', etc.

(iv) the theoretical status of what Jespersen (1965, 308) calls "inarticulate" sentences (cp. "Thanks" and 'The problems of—') and "semi-articulate" sentences (cp. "Thank you" and 'Discusses the problems of—.'), and the relations between such sentences and "articulate" sentences ("I thank you"), both active ('The author discusses—') and passive ('x is discussed'). Jespersen's distinction provides the main subdivisions of the chapter : "Articulate sentences : active" (3.2.1.), "Articulate sentences : passive" (3.2.2.), "Semi-articulate and inarticulate sentences" (3.2.3.).

The third chapter is entitled "Conjunction, relative clauses, and presupposition" (3.3.). The first section is an attempt to explore some marks and factors of identity which are partly common to sentence conjunction and to constructions with relative clauses. The second section (3.2.) carries the discussion a step further and brings in the additional problem of the relatedness between certain noun-phrases with relative clauses and certain lexical items : for instance between "what's known" and "knower" (J 407) (cp. Bach's hypothesis about noun-phrases : Bach 1968). I argue that the observations made in these two sections contradict some of the rules of relative clause formation set up by Langendoen (1970, 141-3) after the Annear Thompson hypothesis, and that they suggest the need for an account of relative clause formation based on Langendoen's concept of "projection" (In 'The whale swallowed Jonah', then, "we would say that the argument the whale receives the specification SWALLOWED JONAH and Jonah, SWALLOWED BY THE WHALE" : Langendoen 1967, 105). In the third section (3.3.3.), on "Asymmetric conjunction and documentation", I argue that asymmetric conjunction is a central feature of the conceptual field of documentation as appears from the equivalence between the word "documentation" and such expressions as

"information storage AND retrieval". Various attested examples
are presented in order to show that, in the conceptual field
of documentation, asymmetric conjunction can be defined in
terms of a set of other, parallel, and partly equivalent,
constructions : for instance, the possibility of having
"searching FOR information retrieval" (J 998), next to
"literature searching AND retrieval" (J 1025). The last section
of the chapter (3.3.4.) examines the relations between asym-
metric conjunction and lexical presupposition as discussed
by Fillmore. It also discusses the problem of asymmetric
conjunction in the field of documentation in relation to Lyons'
concept of implication, which hinges on sentence conjunction
(Lyons 1968, 455 : 'I bought some flowers' "implying" I bought
some tulips AND some roses', etc.), and to the problem of
synecdoche: for instance, "retrieval" standing for the more
usual expression "information storage and retrieval" in "On
retrieval system theory" (J 1044).

On the basis of   observations made in the preceding
chapter, Ch. 3.4. proposes an analysis of asymmetric conjunction
in terms of an extension of case theory (3.4.1.), and returns
to the question of Langendoen's projection and relative clause
formation (3.4.2.). The last two sections present analyses
of various attested examples in terms of the representations
I propose for asymmetric conjunction and cognate constructions
(3.4.3. and 3.4.4.).

The last chapter of my semantic discussion is entitled
"Sem. Complex Instruments" (3.5.). After showing that noun
phrases with such words as "machine" and "system" display
properties typical of Fillmore's Instrumental case (3.5.1.),
I discuss various other examples drawn from the corpus under
investigation which present difficulties for case theory as it
currently stands : for instance, "As machines learn..."

(J 1096) (3.5.2. Some problematic examples). In a third section (3.5.3.), I examine USE and other instrumental predicates, in order to define some sentence types which, to my knowledge, have never been discussed in the literature : consider, for instance, the relation between 'system' and 'techniques' in the sentence "The system (. .) EMPLOYS three established techniques (J 35). The next two sections are devoted respectively to the instrumental (3.5.4.) and locative roles of NP's with 'system' and related words, in connection with Fillmore's proposals to set up multiple case representations (Fillmore 1969a) : compare, for instance, "a system WHEREBY..." and "a system IN WHICH". The last section is entitled "Anaphoric and cataphoric uses of 'system'". In this section, I am chiefly concerned with the role of the word 'system' in discourse structure. In particular, I examine deictic NP's like 'this (the) system', which can occur either in the first sentence of an abstract to refer to sentences that follow, or in the nth sentence of a text without any previous occurrence of the word in the preceding sentences.

The amount of attention I devote to words like 'system' is justified not only by their intuitively obvious importance as probable Instruments in case theory, but also by their role in the conceptual field of documentation, both in the meta-language (the Gardin-Lévy 'concordance') and in the object-language, as is suggested by the following title : "Documentation, information retrieval, and other new techniques" (J 966).

3.1.2. Constituent-structure rules.

This brief presentation of Fillmore's proposals about
constituent-structure rules is based on two main sources, in
addition to Fillmore's publications : Goldin's Spanish case
grammar, and the so-called U.C.L.A. grammar (Stockwell et al.
1968).These, as well as Fillmore's early versions of case
theory are partly outdated, since Fillmore has modified his
thinking on several points (particularly Fillmore 1970b). These
modifications, however, are mostly irrelevant to the main
points of my semantic discussion, with one notable exception :
in the ensuing discussions, I shall follow Fillmore 1970b and
others (cp. Stockwell et al. 1968, Langendoen 1970, 188-190)
in assuming that modality and tense can be represented as
predicates in a 'higher' sentence (which I shall ignore) and
I shall write S for what Fillmore used to symbolize as P.

Fillmore's first version of the constituent-structure
component can bededescribed as follows :
"every Sentence structure (S) consists of a Modality (M)
and a Proposition (P). The Modality constituent includes
the auxiliary and other sentence modifiers. The Proposition
contains the major lexical categories. Specifically, the
Proposition consists of a Verb (V) and one or more case
categories, each of which is divided in turn into a Case Marker
(K) and a Noun Phrase (NP) (...) The exact nature and number
of the case categories remains open to a great deal of investiga-
tion. Fillmore lists eight and says that there are certainly
others. The four cases which are of immediate importance to the
ensuing discussion are the Agentive (A), Dative (D), Instru-
mental (I), and Objective (O)" (Goldin, 5).

Note, in particular, that Fillmore's list of cases is
open-ended. This lack of explicitness is undoubtedly one of
the weak points in Fillmore's proposals. Furthermore, his
theory can be said to rest on four main assumptions.

Assumption 1.

The cases reflect "certain types of judgments human beings
are capable of making about the events that are going on around
them, judgments about such matters as who did it, who it happened
to, and what got changed." (Fillmore 1968a, 24). The case
analysis of a given sentence appeals partly to such conceptual
judgments, and partly to criteria of the type discussed under
Assumption 2. On the basis of these conceptual judgments,
"it is feasible to show that some syntactically different
words are in fact semantically identical", as for instance
like and please (op. cit., 30).

To explain more systematically what the conceptual
judgments involved in case analyses are about, Fillmore has
provided verbal definitions for the cases. The definitions he
gives in the two main published versions of case grammar
with which I am acquainted are tabulated below in (1): I have
changed the order of presentation adopted by Fillmore in order
to display the correspondences that exist, were it only partial-
ly, between the two versions.

(1)

Agentive (A), the case of the typically animate perceived instigator of the action identified by the verb.	Agent (A), the instigator of the event.
Instrumental (I), the case of the inanimate force or object causally involved in the action or state identified by the verb.	Instrument (I), the stimulus or immediate physical cause of an event.

Dative (D), the case of the
animate being affected by the
state or action identified by
the verb.
Locative (L), the case which
identifies the location or
spatial orientation of the
state or action identified
by the verb.

Factitive (F), the case of
the object or being result-
ing from the action or state
identified by the verb, or
understood as a part of the
meaning of the verb.

Objective (O), the semantical-
ly most neutral case, the case
of anything representable by
a noun whose role in the ac-
tion or state identified by
the verb is identified by the
semantic interpretation of
the verb itself; conceivably
the concept should be limited
to things which are affected
by the action or state identi-
fied by the verb (...)
Essive : providing for
sentences of the N be N
type.(Fillmore 1968 a, 24-25
and 84).

Experiencer (E), the entity
which receives or accepts or
experiences or undergoes the
effect of an action (...)
Goal (G) (op. cit., 117)
[no verbal definition is given
but it is contrasted with Source,
which suggests that it is equi-
valent to a former Dative with
a spatial sense]
Source (S), the place from [the
text says, mistakenly, to]
which something is directed.
Result (R), the entity that
comes into existence as a
result of the action.

Counter-Agent (C), the force
or resistance against which the
action is carried out.
Object (O), the entity that
moves or changes or whose
position or existence is in
consideration. (Fillmore
1969a, 116).

In addition to the new definitions tabulated above
in the right-hand column, the second main published version
of case theory (1969a) allows for the possibility of a
given surface constituent having more than one case assigned
to it : for instance, Agent and Source for the subject of an
active sentence with sell.

Moreover, in later unpublished work, Fillmore (1970 b) proposed the following list of cases : A, I, E (in the non-spatial meaning of the former Dative), G (in the spatial meaning of the former Dative), S (in the directional meaning of the former Locative), Place (in the non-directional meaning of the former Locative), Path (as in " he travelled via New York") Time (as in "Summer afternoons are warm"). He suggests that Result can be merged with Goal by generalizing the definition of the latter. He again insists that this list of cases is only tentative and provisional. It seems to me, however, that the most important change which has taken place in the latest unpublished version of case theory is that Fillmore seems to adopt more and more general or abstract definitions for his cases. As a result, many of the original definitions have often become inapplicable. This appears most clearly in the treatment of clauses. In the 1968 version of case theory, Fillmore considered that O was the only case that could be realized as a clause (S) (1968a, 28). He now recognizes that this possibility also exists with other cases, such as I as in "I persuaded John to leave by promising him a reward". Similarly, Binkert (unpublished) suggests that the following sentence contains a sentential Dative (or Goal) : "John devotes a great deal of time to studying rocks". The hypotheses which I shall put forward in 3.4. carry such generalizations of I and Goal a step further.

Assumption 2.

The cases are also assigned, and defined, on the basis of observations about surface grammatical structure. First, Fillmore defines two negative conditions. One is that "only noun phrases representing the same case may be conjoined" (Fillmore 1968a, 22). The other is that "although there can be compound instances of a single case (through noun phrase conjunction), each case relationship occurs only once in a

simple sentence" (op. cit., 21). Next to noun-phrase conjunction, Fillmore seems to recognize other constructions corresponding to compound instances of a single case, such as the compound instruments 'The car's fender' and 'The brevity of your speech' in the following sentences and their paraphrases : 'The car's fender broke the window' ('The car broke the window with its fender'), 'The brevity of your speech impressed us' ('Your speech impressed us with its brevity') (op. cit., 23). In his later, unpublished, work (1970b), Fillmore recognizes yet other possible constructions for compound or complex realizations of a given case : for instance, the Locative in 'He slept in a park on a bench under a tree'.

The positive criteria for assigning a given case concern certain regularities in - or correlations between - various syntactic facts which Fillmore posits to be matters of surface grammar : in particular, subject and object selection, prepositional phrases and the choice of certain prepositions which appear to be typical of certain cases. For instance, 'a hammer' and 'the key' are analyzed as Instruments on the basis of such evidence as provided by the following sentences : 'the key opened the door', 'A hammer broke the window', 'John opened the door WITH the key', 'John broke the window WITH a hammer' (Fillmore 1968a, 22 and 25). It is a matter of further research to show whether or not such criteria will remain applicable under Fillmore's recent, and more abstract, definitions of the cases. Note, for instance, that like 'with the key' and 'with a hammer', 'by promising him a reward' can become a subject when one of the above examples is paraphrased as : 'The promise of a reward persuaded John to leave'.

Assumptions 3 and 4.

These assumptions have not been made explicit by Fillmore himself, but they can be inferred from his treatment of personification in the sentence 'x A hammer broke the glass with a chisel' : "under the personification interpretation", he points out, the sentence "becomes acceptable", and he adds : "what is important to realize is that these interpretations, too, are explainable by reference to exactly the same assumptions appealed to in explaining their 'face value' interpretations" (Fillmore 1968a, 22), that is, presumably to the fact that 'a hammer' when personified is regarded as animate and can therefore be analyzed as Agent. To me, such an analysis rests on two implicit assumptions. One is that 'predicates' with their "case-frame features" (see 3.1.3.) are inserted first, as has been argued by Stockwell et al. 1968. Secondly, considering that Fillmore regards the above sentence as acceptable "under the personification interpretation", considering also that a word like 'hammer' would presumably be assigned an inherent feature [-Animate] according to Chomsky 1965, and that such a feature is incompatible with the definition of Agent, one must conclude that (a) either Fillmore wishes to give up Chomsky's system of unique inherent features, or (b) that the verb selected changes the feature [-Animate] of 'hammer' into [+ Animate]. To my knowledge, Fillmore has never discussed either (a) or (b). What is clear, however, is that his treatment of specifications such as human as the "presuppositions" associated with certain verbs is very similar to Chomsky's use of selectional features, and that in both cases it is necessary to take a stand about (a) and (b). Moreover, there is some overlap between "presuppositions" such as human (see 3.1.3.), and the role of specifications like animate in the definitions of the cases.

As far as the descriptive purposes of this study are
concerned, case theory as it currently stands has serious
deficiencies. Fillmore's versions of the theory as well as the
various studies based on it (e.g. Stockwell et al. 1968,
Langendoen 1970) and the criticisms addressed to it (in
particular, Dougherty 1970) are all based on very simple
sentences of everyday language. There is so to speak no overlap
between the verbs or predicates discussed by Fillmore, his
followers and critics, and the verbs or predicates that need
to be discussed in this study. as they occur in documentation
abstracts and in a corresponding metalanguage.

Goldin's constituent-structure rules, which correspond
to the 1968 version of case grammar are given in (2) :

(2) (a) S --> M P

(b) P --> V (O) (D) (I) (L) (...)

(c) O, D, I, L --> K NP

(d) NP --> Det (N)(S)

(Goldin, 77 and 63)

K in (2) (c) is a Case Marker (Fillmore 1968a, 44-5) : thus,
when K NP serves to rewrite an Agentive, it is mostly realized
as by NP in passive sentences, while this K is realized by
the word order in active sentences. As shown in (2) (b),
one of the main functions of the constituent-structure component
is to define "case-frames" which can be associated with words
represented as verbs or verb-like elements (see 3.1.3.).
In (2) (d), "the linked parentheses mean that at least one of
the two linked elements must be chosen"; rule (2) (d)
"abbreviates three possible constituent structures for Noun
Phrases" :

(a) Det N

(b) Det S

(c) Det NS

(op. cit., 63).

Following various proposals made in the literature, Goldin
suggests that relative clauses are introduced on the basis
of rule (2) (d).

Following the U.C.L.A. grammar, two further rules could
be added to Goldin's rules given in (2). The rules in (3)
would state that a sentence (S) or a noun-phrase (NP) can
consist of a conjunction (C) of two or more sentences (Sx)
or noun-phrases (NPx), respectively. Rule 4 would represent
the hypothesis that nouns can have cases like verbs (Stockwell
et al.'s version of a proposal made in Chomsky 1967) :

(3) (a) S --> C Sx

    (b) NP --> C NPx

(4) N --> N (O) (D) (A) (I) (...)

(cp. Stockwell et al., vol. I, II)

As Goldin points out following Fillmore, a grammar of the
type given in (2) "allows us to provide identical deep
structures for certain sentences with the same meaning and
the same lexical items, but whose surface subjects and objects
are different; for example, sentences 4-6

4   John broke the window with a hammer

5   A hammer broke the window

6   The window broke

What is common to all three of these sentences is the verb
broke, and one noun phrase, the window. The semantic relation-
ship of break to window in all three is presumably the same,
yet window is the object of break in 4 and 5, but the subject
in 6. The relation of break to hammer is the same in 4 and 5,
but hammer is the subject of 5 and the object of a preposition
in 4. Case grammar allows us to state that window is in the
Objective case in all three sentences and that hammer is in
the Instrumental case in both 4 and 5. John, in 4, is in the
Agentive case " (Goldin, 6).

It is easy to think of examples of conjoined sentences
and noun-phrases. Here, I would merely illustrate rule (4)
with examples from Stockwell et al. (1968, 34 and 25) given
in (5) :

(5) (a) John's (= Agentive) _proposal_ (= Noun) of marriage
        (= Objective) to Mary (= Dative).
    (b) the _bleating_ (= Noun) of the sheep (= Agentive).
    (c) the _frightening_ (= Noun) of the sheep (= Dative).

In the rest of the thesis, however, I shall generally
represent structures such as those illustrated in (5) as
S's because, as already pointed out, I have no original
arguments to present for or against the 'lexicalist' hypothesis
adopted by Stockwell et al. in connection with structure like
(5).

To sum up, the basic hypotheses underlying Fillmore's
constituent-structure component are the following :
(a) The _subject-of_ and _object-of_ relations are considered to
    be surface structure relations.
(b) The basic deep structure relations are case relations.
(c) Chomsky's and others' proposed configurational definition
    of deep structure relations is rejected in favour of a
    system which directly labels the arguments of predicates.
(d) Ideally, case theory defines the correlations between
    (a) and (b) by means of transformations.

(•) Until these transformations can be stated, the correlations between observable semantic f ...s (statable by means of cases) and grammatical facts (such as subject, object, and preposition selection) provide the linguist with arguments for upholding his analyses; this, I suggest, is the most interesting aspect of Fillmore's proposals, and throughout my semantic discussion I shall try to present arguments based on such correlations, even when I attempt to go BEYOND Fillmore's framework.

3.1.3. Lexical entries.

The disagreements and uncertainties about the form and place of the lexicon in TG theories of language are well-known. A full discussion of them would require an extremely difficult study which - to my knowledge - has never been attempted, and which will not be attempted here. Similar disagreements exist among linguists working within the theoretical framework set up by Fillmore. For the present purposes, I would simply mention a few proposals made by Fillmore himself. As he does, I wish to mention them as merely examples from which "some suggestions can be gleaned for the design of canonical representations in a lexicon" (Fillmore 1969a, 133). As long as such canonical representations are not available, there can be no question of defining fully explicit lexical insertability conditions - in the sense of lexical insertion as inherited from Chomsky 1965. This qualification should be borne in mind when, following Fillmore, I speak of "lexical entries" in this section.

Before presenting an example of Fillmore's 'lexical entries', however, a few things need to be mentioned. Fillmore assumes that "most of the 'content words' in a language can be characterized in the lexicon in terms of their use as predicates", according to the format inherited from logic : PREDICATE (Argument, Argument, ...) He adds : "I take this to be true of nouns, verbs, adjectives, most adverbs, and also a great many conjunctions" (op. cit., 114). In the paper under review, however, he is especially concerned with verbs. The following are examples given by Fillmore of other catego- ries which he would treat as predicates ; he does not attempt to incorporate these into his case theory - except in one instance (which is in contradiction with the general statement just quoted : KNIFE)

(6) <u>Conjunction</u>

    BECAUSE (x, y)

    (<u>loc. cit.</u>, only example)

(7) <u>Nouns</u>.

    BACHELOR (x) (<u>op. cit.</u>, 123)

    KNIFE

    "use : I of [V O I A]

    where V is 'cut', with the presupposition that O is a

    physical object" (<u>loc. cit.</u>)

(8) <u>Adjectives</u>.

    GOOD (x, ...) (<u>op. cit.</u>, 124)

Another point which needs to be made is that Fillmore
deals with lexical entries in their 'everyday' meaning and
that he does not attempt to define other uses of the words
he analyzes : for instance, although he has been much concerned
with 'verbs of judging', he has never discussed the legal
sense of ACCUSE, and other related verbs. The following table
(9) may serve to summarize his treatment of the predicate
ACCUSE. As already noted, the new case label (Goal) which he
introduces corresponds partly to his former Dative. Furthermore,
the predicate JUDGE, which enters into the definition of others
is left unanalyzed both in the paper under review and in the
paper on "Verbs of Judging" (Fillmore 1969b).

    Lastly, let us note Fillmore's basic distinction between
meaning and presupposition, both of which need to be included
in lexical entries :

> Sentences in natural language are used for asking questions,
> giving commands, making assertions, expressing feeling
> etc. (...) We may identify the presuppositions of a
> sentence as those conditions which must be satisfied
> before the sentence can be used in any of the functions
> just mentioned (Fillmore 1969a, 120). Fillmore (1969b,
> 96) regards some presuppositions as "really preconditi-
> tions" for the use of sentences, thus suggesting a
> logical <u>primacy</u> of presupposition.

As an example, Fillmore gives the sentence <u>Please open</u> <u>the door</u> and he proposes the following test to detect the presuppositional aspect of its semantic structure :

> The test that the existence and specificity of a door and its being in a closed state make up the <u>presuppositions</u> [of the sentence] rather than part of its <u>meaning</u> is that under negation the sentence is used to give quite different instructions, yet the presuppositional conditions are unaffected (<u>loc. cit.</u>).

In the positive as well as in the negative version of the sentence (<u>Please don't open the door</u>), the presuppositions remain the same :

> The presuppositions about the existence and specificity of the door relate to the use of the indefinite article (...) The presupposition about the closed state of the door is a property of the verb OPEN (<u>op. cit.</u>, 121).

(9) <u>Lexical entry for ACCUSE</u>

B.SIC DISTINCTIONS	SYMBOLIZATION	COMMENTS AND EXAMPLES
<u>Arguments</u>	x, y, z	These correspond to "specific roles" corresponding to <u>accuser</u>, <u>person accused</u>, and <u>offense</u>, respectively.
D E E P Str. <u>Cases</u>	Source and Agent = $\underline{x}$ Goal = $\underline{y}$ Object = $\underline{z}$	Cases are abstractions from these roles (op.cit., 115). The subcategorization of verbs in this fashion is Fillmore's method for connecting the constituent-structure rules and the lexicon in his theory. The array of cases associated with a verb is called a case-frame feature.

S U R F A C E	Pre positions (obligatory or optional)	<u>by</u> with <u>x</u> ∅ with <u>y</u> <u>of</u> with <u>z</u>	<u>By</u> will show up in passive sentences, for instance.
stn.	<u>Normal subject</u> <u>Direct object</u>	<u>x</u> <u>y</u>	This holds for active sentences
Z E R O I N G	<u>Zero for</u> <u>indefinite</u>	<u>x</u>	We say, '<u>John was accused</u>' in order to be "non-committal about the identity of the accuser" (cp. '<u>by someone</u>': <u>op. cit.</u>, 119).
	<u>Zero for</u> <u>definite</u>	<u>z</u>	Fillmore would say that '<u>They accused John</u>' "is usable only in a context in which the addressee is in a position to know" what John was accused of (<u>op. cit.</u>).
	<u>Incorporation</u>	<u>y</u>	In the sentence '<u>Ils se contentaient d'accuser</u>' (? '<u>They just accused</u>') "the nature of <u>y</u> is taken as part of the predicate word" (<u>op. cit.</u>, 120).
	<u>Presuppositions</u>	<u>x</u> and <u>y</u> are human <u>z</u> is an activity which need not have taken place, and a serious offense <u>x</u> judges(<u>z</u> is 'bad')	Cp. <u>I criticized John for murdering Jim</u>. ✻ <u>I criticized John, who has never left Ohio, for taking part in the Berkeley riots.</u>
	<u>Meaning</u>	<u>x</u> indicates (<u>y</u> caused <u>z</u>)	This characterization can be described as a "meaning postulate" (cp. G. Lakoff, 1970b).

I wish to end this brief presentation of Fillmore's proposals for lexical entries with a few comments.

First, as appears from the above table, and as Fillmore (1970b) has emphasized, his way of talking of "presuppositions associated with individual lexical items" is to be regarded as an abbreviation for "presupposed belief conditions for the appropriate use of sentences containing the items in question". This position, it seems to me, is very close to - if not identical with - George Lakoff's conception of "relative grammaticality" and of linguistic theory as a device that pairs sentence representations with their "presuppositions about the nature of the world" or "beliefs about the world" (George Lakoff 1969b, for discussion, see my paper : J. Noël 1971). As far as I can see, the concept of pragmatic competence developed in Part I of this study, and the assumed primacy of presupposition are in full agreement with both Fillmore's recent pronouncement and Lakoff's conception. As will be seen presently, the central place granted in this part of my thesis to asymmetric conjunction and my proposal to reduce presupposition to a linguistic representation of asymmetric conjunction (in Ch. 3.4.) are also intended to meet preoccupations similar to those of Fillmore and George Lakoff in their recent thinking.

Second, there is some overlap in Fillmore's theory between the constituent structure subcomponent and the presuppositional aspects of the lexicon. Thus, animateness is both part of the axiomatic definitions of some cases (and crucial in distinguishing between certain cases, such as Agentive and Instrumental) and part of the presuppositional characterization of some predicates : compare, for instance, 'x and y are human' in the above table, and the need to say that 'x is animate' about x in, say, 'x is sleeping'.

Third, and most importantly, let us note that, in Fillmore's theory, the many crucial tasks which a handful of constituent-structure rules are unable to perform are left over to the lexicon. In particular, let us mention : the task of stating which cases are associated (obligatorily or optionally) with which $\underline{V}$'s or $\underline{N}$'s; the conditions under which certain items can be zeroed; the statement of meaning-postulates and presuppositions; perhaps even the setting up of a complete inventory of cases. In fact, it is as though Fillmore felt that the most urgent need in semantics is "for fresh empirical evidence obtained by painstaking study of concrete lexical data" (Weinreich 1966, 473). This is essentially the position I shall hold in this thesis, with the important qualification that I shall attempt to reduce what Fillmore would call lexical presupposition to a form of deep structure representation of asymmetric conjunction inspired by case grammar, and that I shall do this chiefly on the basis of lexical and grammatical facts attested in the Janaske corpus.

3.1.4. Conjunction, discourse, and relative clauses : problems of identity.

In this section, I wish to go beyond "the elementary structure of the "propositional" core of simple sentences" and to review some recent as well as earlier work on sentence conjunction and relative clauses. I wish to argue that the three phenomena under consideration raise a general problem of identity. First, I shall briefly mention the earlier TG treatments of sentence conjunction in order to show in what way they have been deficient and unrevealing from the point of view of semantic representation. I shall do this by first referring to Stockwell et al. whose "Integration of Transforma- tional Theories on English Syntax" can be regarded as reflect- ing the state of the art in 1968. I shall also quote what I believe to be characteristic passages from a paper by Kuroda (1968) on relative clauses, as well as interesting observations made by Ross (1967) on conjunction and relative clause formation. I shall then turn to more recent treatments of sentence conjunction, namely those of Zellig Harris and Robin Lakoff, and to some relatable work by Langendoen and Lyons.

"We know nothing about conjunction". This recent pronounce- ment by Fillmore (1970b) is, I think, fundamentally correct; this will be apparent from the highly tentative nature of Z. Harris's and R. Lakoff's proposals as reviewed below, however novel and interesting these proposals may be. One clear indication of our ignorance about conjunction is that, to my knowledge, no one has ever seriously discussed the question of the insertability conditions for and or other conjunctions : Stockwell et al., for instance, just assign the features + CONJ and + AND to "and" (Stockwell et al. 1968, 967) in their lexicon.

In spite of Katz   and Fodor's hypothesis that "the
sentence break in discourse is simply and-conjunction" or
"the appropriate sentential connective" (Katz and Fodor 1964,
490-1), and in spite of the existence in various studies of a
phrase-structure rule S --> C S x for conjoining sentences
(PS rule 1 in Stockwell et al. 1968, 30; cp. Langendoen 1969,
33) the problems of discourse have until recently never been
recognized as a legitimate pursuit in TG research. Stockwell
et al., for instance, point out from the outset that they have
not attempted to deal with "discourse (sentence connecting)
adverbs such as however, for, so, then, consequently, etc."
(op. cit., 29).

Perhaps one of the reasons for this attitude (next to
theoretical reasons of the type mentioned below in Kuroda's
words) is the one which Bever and Ross invoke in an unpublished
paper on discourse. I am quoting from a lecture by Labov
at the 1970 Linguistic Institute. Bever and Ross are said to
argue that the problem of discourse analysis is linguistically
insoluble. They argue, for instance, that an encyclopedia
would be required to establish the relation between 'the Bible'
and 'Deuteronomy' in the following pair of sentences :
"You should read the Bible. Deuteronomy is one of the ten
great books of the world". The concept of pragmatic competence
was developed in Part I of this·thesis as an alternative
to the kind of conception adopted by Ross and Bever in their
paper.: in pragmatic competence, the relation between the
Bible and Deuteronomy is the same as that between, say, fruit
and apples ("logical inclusion": Leech 1969, 34-35).

To return to Stockwell et al., they are mainly concerned with certain aspects of and conjunction in fairly simple sentences, and they do not attempt to explore any of the semantic problems of conjunction, as the following "survey of problems" suggests:

We are concerned here with what has traditionally been called "coordinating" conjunction. Our primary concern is with structures containing and, but we also attempt to give an account of structures containing but or or. In particular, we shall investigate the structure of sentences like the following, especially (1.a-g) (which must, however, be regarded as a representative sample rather than an exhaustive summary of types) :

(1) (a) John is in the house and Mary is at school.
    (b) John and Bill left.
    (c) I gave the boy both a nickel and a dime.
    (d) I gave the boy a nickel and the girl a dime.
    (e) Emily may be, and everyone agrees that
        Millicent definitely looks, pregnant.
    (f) John and Mary sang and danced respectively.
    (g) Julian ate pears, Jill peaches, and Jake papayas.

(2) (a) (Either) John is playing basketball or his brother
        is jumping on the roof.
    (b) (Either) Jonathan or David played the harp.
    (c) I'll give (either) a nickel to the boy or a dime
        to the girl.

(3) (a) Algernon went home but Nathaniel stayed.
    (b) I gave the boy a nickel but the girl a dime.

In recent treatments of conjunction by generative grammarians, attention has been focused on two major questions : (1) Is there a deep-structure relationship between conjoined sentences (such as (1a)) and other conjoined structures ? (2) If there is such a relationship how many distinct devices (sets of rules or rule schemata) are required to derive these other conjoined structures from conjoined sentences ?

Relevant to the first question is the choice between two possible sources for sentences such as (1.b). First, we might wish to generate the conjoined structure (John and Bill) in this sentence by means of a phrase structure rule like :

(4) NP --> and NP NPx

where (4) represents an infinite schema generating, in the
first instance, structure like :

(5)

```
              NP
           /  |  \
       and   NP₁...   NPₙ
```
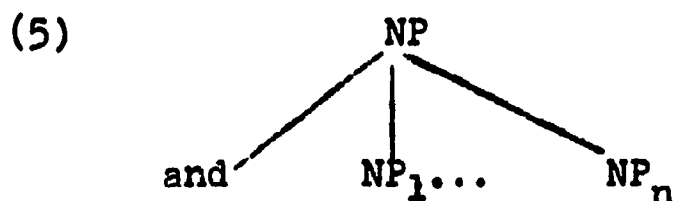
This approach, known as "phrasal conjunction, would
provide for (1.b) a deep structure something like :

(6)

```
          S
        /   \
     MOD     PROP
      |      /  \
    Past    V    NP
      |    /   /  | \
   leave and  NP   NP
              |     |
            John   Bill
```
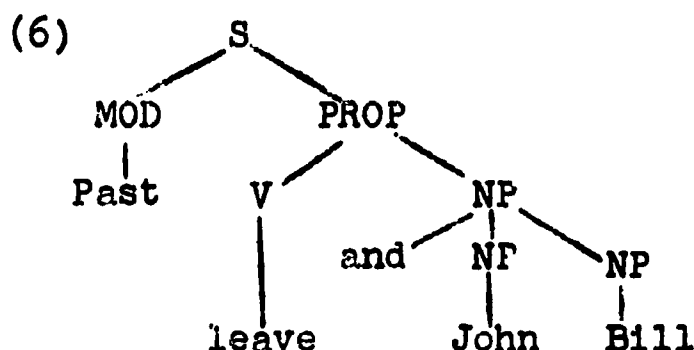
Alternatively, we might wish to say that the deep
structure underlying (1.b) comes from the rule generating
coordinate sentences in the base (PS Rule 1), and is,
roughly :

(7)

```
              S
           /  |  \
        and   S    S
            /___\ /___\
        John left Bill left
```
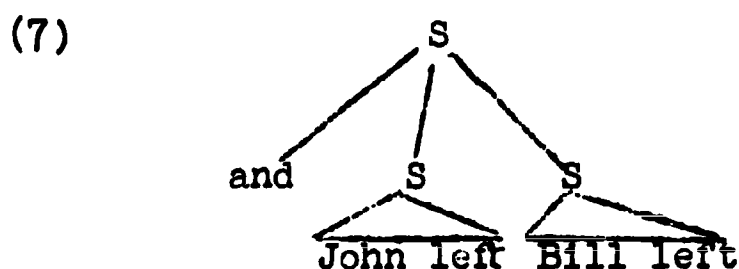
Where a deep structure such as (7) is modified to produce
a surface form such as (1.b), we shall call this process
"derived conjunction". (The process has also been called
"conjunction reduction".)(Stockwell et al. 1968, 321-3).

Stockwell et al. argue that " derived conjunction"
underlies essentially all conjunctions of non-sentences
(ibid., 324).

Another typical attitude in the earlier TG approaches
to conjunction can be seen in a pronouncement by Kuroda
(1968, 1969 reprint). Kuroda notes that there are "semantic
arguments" for relating discourses (such as '<u>something lay
on the table. It was the tissue</u>') to relative clauses (such as
'<u>what lay on the table was the tissue</u>'), but adds that "there
would be no room in the present theoretical scheme of generative
syntax of sentences to say that certain sentences are derived
from certain discourses" (Kuroda, <u>op. cit.</u>, 280).

The only notable exception to this attitude (in the
literature available to me that was published before Stockwell
<u>et al.</u>) is to be found in Ross's thesis (1967). He defines
a "<u>coordinate structure constraint</u>" which blocks relativization
on a single conjunct in a coordinate struct.. e : "In a coordinate
structure, no conjunct may be moved, nor may any element
contained in a conjunct be moved out of that conjunct" (<u>op. cit.</u>,
4.84161). Thus, you cannot move <u>madrigals</u> out of <u>Henry plays the
lute and sings madrigals</u> to form the sentence with a relative
clause x <u>The madrigals which Henry plays the lute and sings
sound lousy</u> (op. cit., 160). Ross then makes the <u>semantically</u>
relevant observation that there are certain exceptions to his
coordinate structure constraint. He considers the sentence
'<u>I went to the store and bought some whisky</u>', and points out
that relativization of '<u>whisky</u>' is possible only if '<u>bought
some whisky</u>' has a purposive relationship to the preceding
clause as in '<u>I went to the store to buy some whisky</u>' : cp.
'<u>Here's the whisky which I went to the store and bought</u>' and
'x <u>Here's the whisky which I went to the store and Mike bought</u>'
(op. cit., 168-9). Ross's conjecture is that "such sentences
do not contain coordinate structures at the time when questions,
relative clauses etc. are formed, but only are converted into
coordinate structures later, or that they never contain
coordinate structures at all". In fact, Ross concludes,

"I know of no other test for coordinate structure than the one provided by the coordinate structure constraint" (op. cit., 170-1).

In (10) (d), I give the only clear example which I have found in the Janaske corpus of a sentence where Ross's test shows that what we have may be no true coordination. Since the sentence in (d) is very complex, I also provide invented examples (a-c) which are similar to (d). I believe that (b) and (c) are acceptable sentences and instances of elements relativized out of the conjuncts of (a) I believe that the attested example given in (d) is parallel to (c)

<div align="center">(10)</div>

(a) The farmer shot the deer with a gun and skinned the animal with a knife.

(b) The gun and the knife with which the farmer shot the deer and skinned it ...

(c) Using a gun and a knife the farmer was able to shoot the deer and skin it.

(d) Using the (....) International System of Notation of Organic Compounds (...), a modified IBM 026 Card Punch with document writing feature and an IBM 866 electric typewriter, it is now possible to assign to each organic compound a single, mathematically unique group of descriptors, and to punch these into a card and retrieve the original cipher in typewritten form at the rate of 600 characters per minute (J 273).

I am acquainted with only two major studies on conjunction since Stockwell et al. and Ross's thesis appeared. One is to be found in Harris 1968 and the other is Robin Lakoff 1969 (1970 version). There are interesting similarities as well as differences between these two, presumably independent, studies. In the ensuing review, I wish to suggest that work along the lines of these two studies could put an end to the ignorance about conjunction which Fillmore deplores.

The following points may serve to sum up Harris's
hypotheses on "conjunctional sequences", i.e. sequences of the
form S_1 C S_2 (Harris 1968, 131-134, except when specified
otherwise) :

(1) Harris shows that conjunctions (C's) impose "certain
    restrictions on word sequence". These restrictions "come
    after those due to sentencehood and before those due to
    discourse". Harris also regards these restrictions as
    "intermediate also in character between the other two".
    In brief, let us say that sentencehood is defined by
    string theory and extensions of it such as the transforma-
    tions, and that Harris regards these as operations on
    sets of sentences which create other sets of sentences that
    preserve the degree of acceptability of those of the first
    set : thus, an ambiguous sentence of the type Frost reads
    well will be transformationally related to two distinct and
    equally acceptable sentence types such as Frost tried to
    read well and One can read Frost well (op. cit., 59 and 55).
    While he grants that "each discourse can be rewritten as
    one long sentence", Harris argues that the restrictions due
    to discourse are "restrictions beyond those that are
    imposed by its component sentences". He describes them
    as "certain factors of modification which operate on
    segments of the first sentence to produce segments of the
    later sentences" (op. cit., 149); as far as I can understand
    him, this happens typically when some segment in one
    sentence gets repeated in a later sentence of a text, but
    co-occurs with material not equivalent to material present
    in the earlier sentence or sentence. Harris contrasts this
    "two-dimensional structure" of discourse with the one-
    dimensional, linear structure SCS ... CS (loc. cit.)
    under discussion here.

This is how Harris further qualifies the possibility of reducing discourses (texts) to a conjunction of their sentences by _and_ or by some other appropriate connective (as Katz and Fodor would put it) :

> The fact that each discourse can be rewritten as one long sentence, if in no other way than by inserting the _and_ of logic between the successive sentences (or, if necessary, between paraphrastic transforms of them) might make one think that the restrictions due to C suffice to account for the restrictions due to discourse. However, this _and_ introduces few if any restrictions, and (...) the restrictions due to discourse are further ones (op. cit., 131).

(ii) In his treatment of conjunction Harris does not concentrate on the differences between conjunctions (e.g. _but_, _because_, _and_, _unless_), since he is primarily interested in stating general constraints which C imposes on the conjoined sentences.

(iii) The first constraint ("minimal difference") is "that certain conjunctions occur between two sentences only if there is at least a particular minimal difference between these" : op. ' x He will write it or he will write it', 'He will write it or she will write it', 'He will write it or he will type it'.

(iv) The second restriction imposed by conjunction is called "word repetition". He establishes this by showing that adding one CS or more that introduce word repetitions raises the acceptability of an SCS sequence : for instance, the "acceptability" of 'The war will start unless he enters the room' is raised in 'Since the invaders threaten war unless the Prince of Cambodia comes to their conference room, the war will start unless he enters the room'.

(v) When a sequence having no word repetition is found acceptable - for instance, The opposition will go underground when the war starts - and when adding a further segment that brings in word repetition - for instance, because open opposition is impossible in wartime - just maintains the

acceptability, Harris argues that the added segments
"have the special property of being known statements of
the dictionary or grammar, or of being common knowledge".
He further suggests that any CS "which adds no information
is zeroable" : for instance, the non-repetitive segments
go underground/open.... impossible in the above example.
Later in his book, Harris posits a category of metalinguistic
sentences to cover all the sentences that "could be
reconstructed by the receiver who understands" a given
sentence (op. cit., 165-6).

Unlike Harris, Robin Lakoff 1969 is partly concerned with
certain differences between the conjunctions she deals with :
essentially and, or and but. Here, however, I shall discuss
the main points of her paper around what she has to say about
and. The main points I wish to mention about her treatment are
the following :

(1) Like Harris, Robin Lakoff argues that conjoined sentences
can be ranked in a "hierarchy of acceptability" (op. cit.,
2). Robin Lakoff uses the term 'presupposition' in the
broad sense of knowledge assumed or presupposed by a given
sentence (cp. George Lakoff 1969b and others), and the term
'deduction' in the informal sense of inferences based on
such presuppositions. Both her notions of presupposition
and deduction correspond to Harris's notion of "common
knowledge" (Harris 1968, 134). Robin Lakoff argues that
the degrees of acceptability of conjoined sentences are a
matter of presuppositions and deductions : 'certain ones,
she says, are universal, or at least assumed to be wide
spread within a culture; others are confined to a subclass,
or are idiosyncratic with the speaker" (op. cit., 55).
Thus, in the sentence 'Felix is a cat and he has four paws'
the first part involves "presuppositions arising from
knowledge of the world" "that make the second part empty".

The oddity of 'John has a Ph. D. in linguistics and he can read and write' is presumably a matter of more specialized knowledge: " to say of someone that he has a Ph. D. involves a chain of deduction, part of which contains the idea that he has completed a course of study, therefore, presumably, has been forced to do reading and writing as a prerequisite for earning the degree" (op. cit., 15). Like in the preceding sentence about 'Felix' , the second conjunct is part of the presuppositions associated with the first. In other words, the second conjunct is tautological like that in x John is rich and he has 80 million dollars (op. cit.,5). Among the sentences which require "even more specialized knowledge" to be fully acceptable - knowledge of Chinese cuisine, and, for a non-American, of American life - Robin Lakoff also discusses 'John wants to make Peking Duck, and I know that A and P is having a sale on hoisin sauce'. In this connection, she makes the interesting po¹nt that such a sentence is ambiguous in a sense so far not discussed in the literature : its ambiguity is a matter of what "the hearer" knows about the relationship between Peking Duck and hoisin sauce. Robin Lakoff suggests to describe this ambiguity as "situational or contextual, as opposed to structural ambiguity" (op. cit., 9 and footnote 2). Her concept of contextual ambiguity is the only theoretical concept I have found in the linguistic literature that corresponds to the multiplicity of meta-languages as discussed in Part I of this thesis.

(ii) With respect to (1), the differences between Harris and Robin Lakoff can be explained as follows. Harris proposes the single 'grammatical' and one-dimensional mechanism of adding CS's that may or may not increase the acceptability of a given string of conjoined sentences. To understand better what Robin Lakoff proposes, one needs to remember the conception of 'grammar' (linguistic theory) put forward

by George Lakoff : "a grammar characterizes the relation
between surface forms [surface structures], logical forms
[semantic representations], and classes of contexts
[essentially, presuppositions and deductions] by generating
an infinite class of pairs of derivations [relating semantic
representations and surface structures] and classes of
contexts" (George Lakoff 1971, 14). Corresponding to this
conception is Robin Lakoff's suggestion that what belongs
to the presuppositions and deductions associated with a
sentence   should be _separated_ from the actual (:structural')
semantic representation - as will appear from examples
given below in (111). Moreover, Robin Lakoff recognizes
explicitly that there are LOGICAL mechanisms involved in
what makes conjoined sentences acceptable. This is particu-
larly clear when she discusses the sentence _George likes_
Peking Duck, but who wouldn't ? :"the.rhetorical question
who wouldn't ? really means everyone does" and this, she
points out, suggests that '_George_' of the first conjunct
must be "logically included" in _everyone_ of the second
(_op. cit._, 39-40). Similarly, Robin Lakoff suggests that
the very odd sentence '_Six men can fit in the back seat of a_
Ford and I wrote my grandmother a letter yesterday' may be
interpretable because "the generic comes first" and establish-
es "a true-for-all-time condition for the causation of the
second, punctual conjunct". It is then interpretable on the
basis of a set of assumptions or presuppositions, like the
following : "I made a bet sometime previous to yesterday
that, if six men could fit in the back of a Ford, I would
write my grandmother a letter. It was duly proved that six
men can do this, so yesterday I did what I promised".
(_op. cit._, 21). Harris, as far as I can see, does not
recognize, or allow for the possibility of recognizing,
the logical nature of such facts.

(111) "For all conjunction, a common topic is necessary. This
may be overtly present in the superficial structure of the
sentence, or may be derivable by more or less complex
combinations of presuppositions + deductions. In any
event, to account for the kinds of superficial structures
that are amenable to conjunction, we need to incorporate
the concepts of presuppositions in our grammar" (op. cit.,
55). Here, the logico-semantic notion of "common topic"
corresponds to Harris's word-repetition constraint and
CS's having the form of definitions representing "common
knowledge". What Robin Lakoff describes as common topic
is best explained by means of a few examples. Robin Lakoff
begins by noting that "usually" the common topic is NOT
"overtly present and identifiable in the sentences"; in fact,
when this is the case "conjunction reduction or gapping"
usually occurs, deleting the identical constituents and
reducing a sentence John eats apples and John eats pears
to John eats apples and pears (op. cit., 6, 5 and 2).
In most cases, "the hearer must supply, from his experience
or knowledge of the world, or from prior discourse he has
participated in, additional facts that link something in one
conjunct to something in the second" (op. cit., 6-7), and
this is done by making presuppositions and performing
deductions. Consider the sentence already quoted above :
'John wants to make Peking Duck, and I know that the A and
P is having a sale on hoisin sauce'. At another place, she
insists that the common topic is "based on semantic rather
than on purely lexical similarity" (op. cit., 28) - a view
which is apparently in opposition with Harris's account
in terms of "word repetition". One of the set of presup-
positions given by Robin Lakoff for the right-hand conjunct
of the sentence about Peking Duck is :

1. (presupposition :) Hoisin sauce is the usual accompaniment to Peking Duck.
2. If there is a sale, it is a good opportunity to buy what is on sale.
3. Therefore, now would be a good time to make Peking Duck.
4. To make Peking Duck = to make Peking Duck.

In connection with this - in particular 4 - she speaks of "deduced identity" as the basis for the identity of topic in the conjuncts. The difference with Harris may not be as clear-cut as may seem at first - except for the appeal to logic in R. Lakoff's discussion. Note that the above analysis could be coverted into an SCS sequence creating word repetition, as Harris would suggest. For instance :

> John wants to make Peking Duck, and I know that the A and P is^{a}store and is having a sale on hoisin sauce; and, since Hoisin sauce is the usual accompaniment to Peking Duck and since a sale is a good opportunity to buy what is on sale, now would be a good opportunity for John to buy hoisin sauce and to make Peking Duck, etc.

Irrespective of the validity of the details of both analyses, however, it seems to me that Robin Lakoff's insistence on the logical character of presuppositions and deductions (for instance, some inclusion relation holding between John's situation and what is 'usually' done) is a necessary ingredient of the analysis of conjunction, and one which Harris fails to capture.

(iv) Unlike Harris, R. Lakoff makes no difference between sentences and discourses. At one place she calls the following a "discourse" (op. cit., 22) and at other places "a sentence" :

(11)

> Well, the story is as follows : The police came in, and everyone swallowed their cigarettes, and Bill choked on his, and they had to take him to the hospital, and his mother just about went frantic when she heard, and I had to placate her by lending her my copy of Portnoy's Complaint.

(v) At various places, Robin Lakoff traces a very tentative but most interesting connection between some uses of conjunction and Ross's hypothesis about implicit performatives (Ross 1970). For instance, she notes that the sentence with but given in (a) can be paraphrased as in (b) - under the assumption that 'George is a linguist' and that 'Peking Duck is Chinese food' :

(a) George likes Peking Duck, but all linguists are fond of Chinese food.

(b) I say to you that George likes Peking Duck, but I really don't have to say this, because all linguists are fond of Chinese food. (op. cit., 39 and 42).

(vi) The logical nature of the common topic constraint appears most clearly in the distinction which - unlike Harris - R. Lakoff makes between symmetric and asymmetric conjunction.

Symmetric conjunction can be characterized by the fact that it is reversible without any change in meaning : 'Fred is chasing the aardvark and Mary is eating toast'. But like in the asymmetric type, "a common topic is necessary if the conjunction is to be acceptable" (op. cit., 24). For 'John has a yacht and Bill has a lovely home in Scarsdale', for instance, a possible presupposition one might regard as the common topic is "Owning a yacht is an example of conspicuous consumption, and having a lovely home in Scarsdale is too" (op. cit., 25).

Asymmetric 'and' is "equivalent to and then, in either a temporal or a causal sense" while symmetric and simply links two or more sentences without imposing an order of priority on the sentences it links. The discourse of (11) is an example of the asymmetric type. With this type of and, R. Lakoff argues, "the first member of the pair is presupposed for the second to be meaningful" (op. cit., 22).

If one attempts to object to (10) by saying '_But the police never came in_', "the result is bizarre, and renders the whole discourse somehow nonsensical" (_op. cit._, 22). Moreover, the sentence '_Everyone swallowed their cigarettes and the police came into the room_' is either a "symmetric conjunction, or it makes very different assumptions about causality than does" '_The police came into the room and everyone swallowed their cigarettes_' (_op. cit._, 20). In asymmetric conjunction, "the whole is, in a sense, greater than the sum of its parts : besides the stringing together of a number of facts (the sum of its parts), the whole also includes the idea that each one led up to the next, and that one would not have occurred, or would not be true, except that the ones that preceded it were true. Taken out of context, the conjuncts would not retain this causal or implicational link so that part of the meaning of the sentence would be lost". (_op. cit._, 23). Robin Lakoff conjectures that with asymmetric conjunction "it is the two conjoined sentences as wholes that participate in the relationship" marked by _and_, and that the common topic is "the causality or temporal priority" (_op. cit._, 26).

Obviously, Robin Lakoff's pronouncement that "taken out of context" part of the meaning of the conjoined sentences would be lost needs some qualifications as I shall try to show later. This can already be seen if we examine one of R. Lakoff's examples just quoted : _I had to placate her_, for instance, necessarily presupposes that the female person in question is _excited_, _frantic_, or the like, this holds irrespective of the context and can be regarded as something that has to do with _placate_, _need placating_ ,etc. This is a lexical presupposition, of the type discussed by Fillmore. To quote one of Fillmore's clearest

examples (Fillmore 1968c, 52) the fact of stating in isola-
tion a sentence with _borrow_ does not affect its presuppo-
sitional meaning. Compare (12) (a) and (b) :

(12)

(a) John had a book and Bill borrowed it from him.
(b) Bill borrowed a book from John.

(12)(b) carries the presupposition that <u>John had a book</u>
(before the time of borrowing) whether or not this sentence
is uttered before (b) and conjoined with it. Or, to take
another example, if somebody <u>concludes</u> that such and
such is the case, we need not be told that something
came before and that the one who draws the conclusion
.bases it on that previous event. Similarly,
when somebody <u>answers</u> somebody else something must neces-
sarily have been said before the answering. In fact,
with verbs such as <u>borrow, conclude, buy/sell,</u> <u>agree</u>
(not to mention adverbs like <u>again</u>) it is particularly
obvious that the very notion of a simple sentence in
isolation just collapses. Lexicographers show awareness
of this in their examples: for instance, <u>I asked him to</u>
<u>help me and he agreed</u> (Hornby's entry for <u>agree</u> : Hornby
1963). What this goes to show is that statements of
identity will have to incorporate facts of the type
discussed by Fillmore around the concept of presupposition.
My arguments here are similar to (unfortunately still
informal) proposals put forward by some logicians. Thus,
in an attempt to reconcile the <u>atomistic</u> position of
the semantic approaches based on Chomsky with the <u>holistic</u>
positions of structural semanticists. Gochet 1970a puts
forward a notion of "double context", after a suggestion
made by Rivetti Barbo to distinguish between <u>the context</u>
<u>provided by the sentence</u>, and <u>the context provided by the</u>
<u>lexicon.</u> Reference would be based on the former and <u>sense</u>
on the latter. As Gochet points out, 'sentence context' can
be responsible for semantic <u>redundancy</u> (as in <u>The bird flies</u>)

or _selection_ (as in _He drank the stuff_) but never for a
qualitative change in meaning. Gochet also quotes Rivetti
Barbo who writes : "if meaning were to depend on the
context provided by the sentence, it would be impossible
to understand new sentences that have never been heard
before" (Gochet 1970a,229-230).

(vi) As a tentative conclusion to her discussion on _and_
R. Lakoff advances the following hypothesis. She notes
that 'A and B are both examples of conspicuous consumption'
allows the further deduction, 'A and B are similar' (i.e.
a sentence with a symmetric predicate, which can take
any number of conjuncts, as in _Bill, John and Harry are
similar_) whereas asymmetric conjunction can be reduced
to a sentence with an asymmetric predicate such as _A follows
B_ (_≠ B follows A, x A follows B follows C_). "Then, perhaps,
what ¬ne means by symmetric conjunction, and common topic
in connection with it, is 'possibility of reduction by
presupposition and deduction to an underlying symmetric
predicate', opposed by asymmetric conjunction involving
reduction to an asymmetric predicate" (_op.cit._, 27).

Neither Harris nor Robin Lakoff include the question
of _relative_ clauses in their discussions of conjunction. I
believe this to be regrettable and will try to fill the gap
here on the basis of some of Langendoen's work. The reasons
for discussing conjunction and relative clauses together are,
I think, fairly obvious. For one thing, there exist paraphrase
relations between many conjoined sentences and sentences with
relative clauses. These have been mentioned above in connection
with a paper by Kuroda and have led Annear Thompson and
Langendoen (among others) to posit that sentences with relative
clauses are transformationally derived from conjoined sentences.

But an even more important reason for linking the two
problems is that for both types of structures it is necessary
to state that some _identity_ holds between constituents in two
or more sentences. Thus, in all the traditional accounts of
relative clause formation based on Chomsky's formulations
an identity condition needs to be stated between an NP in the
matrix sentence and an NP in the embedded sentence in order to
transform the phrase-marker into a sentence with a relative
clause : for instance, the identity involving _the man_ in "_the
man_ (_the man_ had been fired) returned to work" (cp. Chomsky
1965, 137 and 145). The same is true of the Annear Thompson
proposals to derive sentences with a relative clause by means
of optional transformations from pairs of conjoined sentences
(Annear Thompson 1970a and b· Langendoen 1970, 141-5).

Now, as Langendoen has shown, the identity problem appears
in a new light in conjoined sentences or in discourses as soon
as one considers the effect of what he calls _projection._
Thus, in _The whale swallowed Jonah_ "the argument _the whale_
receives the specification SWALLOWED JONAH and _Jonah,_
SWALLOWED BY THE WHALE", and, in a discourse, the following
sentence is incompatible with the first : _However, Jonah managed
to escape being swallowed_ (Langendoen 1967, 105). If this is
correct, as I believe it is, at least part of the identities
in discourses will have to be stated in a form closely related
to that of relative clauses, as in _The whale swallowed Jonah
but Jonah (who had been) SWALLOWED BY THE WHALE escaped._
The problem is also related to that of pronouns (_he_ instead of
Jonah) and deictics (for instance, _the man_ for _Jonah_).
Thus, it seems to me that projection would have to be taken
into account in rules of relative clause formation - including
Langendoen's own formulation of these rules.

To close this review of what I believe to be important contributions to the study of conjunctions, it is necessary to refer again briefly to Lyons's work on implication and hyponymy (Lyons 1968, 453-5). Lyons suggests that a sentence 'I bought some flowers' implies EITHER a 'disjunction' 'I bought some tulips OR I bought some roses OR I bought some violets...' OR conjunctions 'I bought some roses AND some tulips', OR 'I bought some violets AND some tulips', etc. This could no doubt be extended to other parts-of-speech categories than nouns, as appears from many dictionary definitions : for instance mean 'low, bad, unkind, OR dishonest in small ways', fetch 'go to another place AND bring a thing from it' (West 1965). I shall return to these questions later. Here I would merely say this : no matter how problematic Lyons's account of implication may be, it suggests that his concept of implication is part of the broader problem of conjunction between sentences, and of the identity which Harris and Robin Lakoff posit to hold between conjoined sentences.

To conclude, the following points may be made; the underlying assumptions and claims will be examined in chapter 3.3.

(i) Harris and Robin Lakoff have suggested that sentence conjunction imposes some kind of identity constraint on the conjoined sentences : "word repetition" or zeroable CS's expressing "common knowledge", according to Harris, and a "common topic" based on "presupposition" and "deduction" according to R. Lakoff.

(ii) If we follow R. Lakoff's suggestion, presupposition and deduction have certain logical properties which need to be formulated in a linguistic theory.

(iii) R. Lakoff's distinction between symmetric and asymmetric conjunction is an important one. In particular, it raises the problem of whether asymmetric AND is basic or derived,

and of how to state its relations with superficially
distinct items - for instance, functions of purpose
(Ross 1967, 168-9), asymmetric verbs such as <u>cause</u>, <u>follow</u>,
etc.

(iv) The following point is implicit in (iii) as well as in
Harris's contention that intersentence <u>and</u> in discourse
is "logical" <u>and</u>. It seems that one essential value of
asymmetric <u>and</u>. as compared to other predicates, is that
it blurs, obscures or neutralizes relations which are
made explicit by other predicates or by various other
linguistic means (such as time adverbials). This can be
shown in various ways . Jespersen, for instance, made
similar observations in his discussions of nexus-
substantives and rank-shifting ;compare the following :

(13)

Sentences with nexus-substantives	Sentences connected with <u>and</u>, etc.
(a) The Doctor's extremely quick ARRIVAL AND uncommonly careful EXAMINATION of the patient <u>BROUGHT</u> <u>ABOUT</u> her very speedy RECOVERY.	(a') The Doctor ARRIVED extremely quickly <u>AND</u> EXAMINED the patient very carefully; she RECOVERED very speedily.
(b) Close upon his RESIGNATION <u>FOLLOWED</u> his last illness and death	(b') He RESIGNED <u>AND</u> shortly after he BECAME ILL <u>AND</u> DIED [my paraphrase]

(Jespersen 1965, 137-8)

A similar situation has been illustrated as follows in a recent study of "interclausal relations" in a Philippine language (Inibaloi) :

> Suppose, for example, a situation in which John leaves the apartment and goes downtown exactly seven minutes by the clock before Mary, his wife, comes home. John may report this event the next day in the words, 'I went downtown and Mary came home'. If he chooses this wording, he disavows interest in temporal considerations. On the other hand, John may find himself in a situation where it is of crucial importance to establish the succession of events. He may therefore say, 'I went downtown and then Mary came home'. In still other circumstances John may want to emphasize that the two events were not far apart and that succession between them is of no importance. In this case he may say 'About the time I went downtown Mary came home'(...) This analysis of John's three ways of reporting the same situation does not reduce the distinctive conjunctions to synonyms, but preserves their essentially different thrust, while at the same time it recognizes the option of the speaker to conceptualize the same situation differently.
> (Ballard et al., 1971, 72).

(v) A further question which arises from the studies reviewed above concerns the relations between conjoined sentences, texts or discourses and relative clauses. In particular it is necessary to test the hypotheses about the possibility of reducing all discourses or texts to a conjunction of their sentences by and or some other appropriate connective, and to compare these hypotheses with the Annear-Langendoen proposals about relative clause formation, and Langendoen's notion of projection.

(vi) It may be necessary to examine facts of the type discussed by Fillmore around his notion of presupposition and of the type discussed by Lyons around implication and hyponymy to capture some aspects of the identity problem in sentence conjunction and relatable constructions.

(vii) It seems that R. Lakoff accepts the derived conjunction hypothesis about noun-phrase conjunction (and presumably verb-phrase conjunction) for instances of symmetric conjunction.

It remains to be seen whether the hypothesis can apply
to asymmetric conjunction.

(viii) Lastly, let us mention the problem of relating the
implicit performative hypothesis with the Harris/Lakoff
hypotheses about conjunction.

3.2. 'A DISCUSSION OF ...'

3.2.1.1. Articulate sentences : active.

In this chapter, I wish to examine sentences of the type 'The author discusses -', and relatable constructions, such as 'A discussion of-' as they stand, for instance, at the beginning of an abstract. Following Jespersen (1965, 308), I shall distinguish between "articulate sentences" ("I thank you"), "semi-articulate sentences" ("Thank you") and "inarticulate sentences" ("Thanks"). Among the articulate sentences, I shall discuss successively active constructions (in this section, and the next : 3.2.1.2.) and passive constructions (in 3.2.2.). Semi-articulate and inarticulate sentences will be dealt with in section 3.2.3. My inventory of the expressions examined in this chapter is based on the assumption that they convey no information about the subject or topic of the abstract in which they appear but rather serve to introduce this topic. The purpose of what follows is to examine on what theoretical foundations such a working hypothesis could be based

The following points will not be discussed, in spite of their possible interest :

(i) The problem of tenses. The most frequent tense is the present, as in 'The author discusses-'. Following Zandvoort, it may be called a "neutral present", "used when no particular time is thought of" (Zandvoort 1966, 59). Occurrences of the present perfect are far less frequent : "The authors have included -" (J 975). So is the preterit, as in "On the basis of these experiments we concluded that-" (J 150).

(ii) Various problems which could perhaps be dealt with by means of transformations. Compare : "A world-wide survey..." (J 103) and "A survey of the world-wide..."(J 992).

(111) References to the <u>reader</u> of the abstracts . I found only a few examples, as in "...may guide the <u>interested reader</u>" (J 705), "In the second chapter the authors give <u>us</u>-"(J 38).

(iv) <u>The question of 'implicit performatives'</u>. Following Harris (1968, 79 f) and Ross (1970), all declarative sentences can be analyzed as object clauses depending on a declarative main verb : e.g. '<u>I declare to you that -</u>'. In the abstracts under investigation, the <u>I</u> would stand for the abstractor, or for the author in the case of an author's abstract. In the Janaske corpus, I have found a small number of sentences which may require reference to the implicit performative hypothesis - although I have no idea of how to explain them. Some are passive sentences, which are discussed below (3.2.2.). The others are essentially expressions which reflect the <u>abstractor</u>'s opinion about what he reports. Consider the underlined words in the following sentences, in particular some adjectives and adverbs :

(1)

(a) <u>Of special interest is chapter 6</u> (...) which is concerned with - (J 524).

(b) <u>Of particular interest is the lucid</u> description of - (J 687).

(c) The presentation is in tabular form which offers <u>easy</u> comparison of - (J 499).

(d) The engineer has been joined in the search by physicists, biologists, neurologists, psychologists, and linguists, <u>not to mention</u> mathematicians. (J 258).

(e) The <u>interesting</u> observation is made that - (J 507).

(f) Included are <u>extensive</u> tables (...) and a <u>generous</u> list of references <u>vital</u> to <u>seemingly</u> every aspect of this field (J 715).

(g) Contends that... <u>To make a comment on these comments,</u> ... (J 1104).

(v) <u>Expressions for 'AUTHOR' and 'TEXT'</u>. I shall also ignore
the problems concerning what can stand in the place of
'author' and 'paper' in sentences of the type 'In this
<u>paper</u> (the <u>author</u>)discusses -'. In the Janaske corpus,
I have found that what may be referred to as the 'AUTHOR'
role can be played not only by words like '<u>author</u>' or
'<u>writer</u>', but also by the name of the person in question
(e.g. 'Dr. So-and-So') or by noun-phrases referring to the
person's function : for instance, "The president of Stanford
Research Institute discusses- " (J 149). The 'TEXT' role
can be played by various words as are listed in (2).
The words of the (a) list have as a determiner either the
demonstrative <u>this</u>, or <u>the</u> (<u>present</u>) with a demonstrative
value. Very often, this role is played by just the
demonstrative pronoun <u>this</u> (c) as in "<u>This reviews-</u>"(J 2).

(2) 'TEXT'

(a) 'WHOLE'	(b) 'PART'	(c) DEICTIC ONLY
article	appendix	this
bibliography	bibliography	
book	chapter	
compilation	glossary	
dictionary	part	
discussion	photographs	
edition	supplement	
glossary	talk	
issue		
letter		
manual		
material		
narrative		
note		
(these pages)		
publication		
report		
review		
study		
survey		
volume		

Here I shall not discuss these various points in order
to concentrate on two questions : I shall delineate a case
analysis of sentences of the type 'The author discusses - ';
and examine the possibility of relating some of the sentence
types under discussion to each other by means of rules of
implication.

Following Fillmore, I shall argue that "the subject
relation is a result of neutralization of other deeper
grammatical relations" (Hofmann 1969, 12) and I shall attempt
to determine what these deeper grammatical relations could
be in the active sentences under discussion in this section.
"The ideal technique" for doing this "is to find a systematic
pairing of sentences which are paraphrases and which contain
the same lexical items, in such a way that one of each pair
has one nominal (or "NP") as subject while the other has
another nominal subject" (loc. cit.). Applying this technique,
we observe that in active sentences, the verb 'discuss'
can have two clearly distinct types of subjects, which can
be called 'AUTHOR' and 'TEXT' respectively. Moreover, when
'AUTHOR' is the subject, 'TEXT' can be used in prepositional
phrases introduced by in. Compare (3) (a), (b), and (c) :

(3)

(a) The author discusses ——
(b) This paper discusses ——
(c) In this paper, the author discusses ——

It would be worth investigating to what extent various
words other than those listed in (3) behave like 'paper'
in (3) : 'grammar', 'novel', 'rule', 'signpost', t-mb-stone,
etc.

Extending the observations made in (3) to the whole of
the Janaske corpus Icame up with a fairly large list of verbs
all appearing in active sentences; this list is given in (4).
Whatever the semantic as well as syntactic differences between
these verbs may be, I suggest that they all accept the three
types of constructions just illustrated above in (3) with
'discuss'. The verbs which were actually found as accepting
both 'AUTHOR' and 'TEXT' as subject are underlined in list (4) :

(4)

'AUTHOR' or 'TEXT' or 'THIS' AS SUBJECT

adds that
admits that
advocates
argues that
analyzes
asserts that
assumes that
begins with
cautions against
cites
characterizes -as-
chronicles
claim
compares
is concerned with
concludes that/with
conclusion is that
considers NP/that
contends that
covers
criticizes
deals with
defines
demonstrates
details
describes
develops (X)
devises (x)
discusses
(re)emphasizes
ends with
estimates that
evaluates
explains

mentions
names
notes
offers
outlines
points out that/NP
points up
predicts
presents
provides
proposes that/Ving
recognizes that
recommends
recounts NP
refers the reader to
refers to
reflects
reports
reports on
reviews
sees..as
shows
states that/-to + V
stresses (the fact) that
suggests
summarizes
surveys
tells of
touches (up)on
traces (history)
treats
turns to
unearth
urges that/-to + V
views

explores   warns against/that
favors   writes that
forecasts
gives (us)
illustrates
includes
indicates that /NP
insists that
lists

(∝.) "The author develops a communication system..." (J 606).
(∝) "The paper devises a method for ..." (J 908).

    The verbs of the above list can now be opposed to other
attested verbs ('or cognate expressions) that have a different
syntactic behaviour : some seem to accept ONLY 'AUTHOR' AS
SUBJECT (5), and others ONLY 'TEXT' OR 'THIS' (6). In addition,
I have found one expression with 'TEXT' as prepositional phrase:
"___ is the subject of this paper" (J 622) (cp. 'The subject
of this paper is ___': not attested).

                                (5)
                    ONLY 'AUTHOR' AS SUBJECT

        (a) believes
            is convinced of/that
            feels that
            is hopeful that
            lives with
            is of opinion that
            is pleased with
            is satisfied with

        (b) it became clear to ── that

        (c) ──'s belief is that
            ──'s concern is with

                                (6)
                    ONLY 'TEXT' or 'THIS' AS SUBJECT

        (a) amounts to
            is concentrated on
            contains
            is designed to
            is devoted to
            is divided into (-parts:)
            is in (-parts)
            is like nothing so much as

```
          is limited to
          occupies (pp...)
          is on
          relates to
     (b)  is an account of
          is an analysis of
          is an appeal to (-to V)
          is an attempt to
          is a bibliography
          is a (reference) book for
          is a description of
          is a dictionary of
          is a discussion of
          is an evaluation of
          is an examination of
          is an explanation of
          is an exploration of
          is a guide to
          is an illustration of
          is a list of
          is a note on
          is an outline of
          is a plea for/to (-to V)
          is a report on/of
          is a review of
          is a study of
          is a summary of
          is a survey of
```

In what follows I shall concentrate on just a few verbs around the following points :

(i) A Locative analysis of 'TEXT';

(ii) An Experiencer analysis of 'AUTHOR';

(iii) Arguments for a topic case;

(iv) Prepositions as predicates versus 'essive' analyses.

(i) A Locative analysis of 'TEXT'.

The strongest argument for a locative analysis of 'TEXT' in sentences with verbs such as 'discuss' (list 4) is the following. If Fillmore is correct in assigning the locative case to the expressions underlined in (7) (a), then those underlined in (7) (b) must also be analyzed as locative :

(7)

	(a)	(b)
(1)	"IT is windy <u>in Chicago</u>".	"IT says <u>in the Bible</u> (that)..." (Jespersen 1966, 238; Zandvoort 1966, 59; not in Janaske).
(ii)	"<u>Chicago</u> is windy". (Fillmore 1968a, 25)	<u>This paper</u> discusses —
(iii)		<u>In this paper</u>, the author discusses—

Besides, I have no attested examples of passives like the following : 'x is discussed <u>by</u> this paper', 'x has been included <u>by</u> the paper'. In addition, an examination of the Janaske corpus shows that both in the active sentences discussed here and in the passive sentences to be discussed in section 3.2.2., only the preposition '<u>in</u>' shows up, next to clearly locative adverbs in the passive constructions: "___ is HERE presented" (J 486), "___ are reported HEREIN" (J 298). The only exception which I have found may be attributed to special properties of the verb <u>guide</u> : "The description is rather short, but the reader is guided to more detailed information on the subject <u>through</u> the bibliography provided <u>with</u> the article" (J 928 : the '<u>with</u>' is presumably to be regarded as a comitative rather than an instrumental '<u>with</u>').

The only other case that could serve to analyze 'TEXT' is presumably the Instrumental case. This hypothesis, however, must be rejected for two reasons. First, the possibly instrumental expressions which are associated in the Janaske corpus with verbs such as '<u>discuss</u>' are different from 'paper', etc.: "<u>On the premise that</u>___ the writer discusses ___" (J 289), "In addition to defining ___ <u>by using a simple illustration of</u>___, the author shows—" (J 82), "<u>The concept of quantity is used to describe the statistical correlation</u> ..." (J 76; cp. 'x The article is used to describe').

Second, there is one range of examples in which the
Instrumental analysis <u>contrasts</u> with the Locative one.
These are sentences with verb such as '<u>prove</u>', '<u>show</u>',
'<u>indicate</u>', '<u>suggest</u>', etc. to which Fillmore 1970b would
assign an Instrumental case. Thus Fillmore would, I think
rightly, analyze the first that-clause as an Instrumental
in "<u>That Harry is still living with his mother</u> PROVES that he
is a bad marriage risk" (Fillmore 1969a, 121), which could
be simplified into "<u>This</u> PROVES that___". Now, note that
"<u>This</u> PROVES that____", has a different meaning if '<u>this</u>'
refers to some 'TEXT' as in 'This paper proves that—'.
The following, similar examples are drawn from the Janaske
corpus; I suggest to interpret '<u>this</u>' as a Locative in (a),
and '<u>they</u>' as an Instrument in (b); and that a similar
contrast is present in (9) :

(8)

(a) A discussion and evaluation of the suggestion by ___
    for "A Concise Form for Scientific Literature
    Citations". The suggestion centers around the use
    of four letters for the title of each journal.
    <u>This suggests that</u> the confusion resulting from the
    use of these abbreviations may exceed any possible
    benefits from the space saved. (J 372).

(b) While many of the findings are still of a preliminary
    nature, <u>they show</u> the need for further development...
    (J 690).

(9)

(a) <u>Neither method</u> TELLS much about scientists' use of
    current literature (J 1051).

(b) <u>The authors</u> TELL how the procedures and cards (...)
    would be changed if they were to start again (J 671).

(c) <u>The article</u> TELLS of the techniques and devices under
    investigation (J 1063).

I can only think of verbs such as 'illustrate' and
'exemplify' with which the contrast between the Instrumental
and the Locative analysis is not clear, as in the following
sentences :

(10)

(a) A chart illustrates this problem.
(b) The author uses a chart to illustrate this problem.
(c) The author illustrates this problem with a chart.
(d) The problem is illustrated in a chart.

I am unable to provide evidence for or against a locative
analysis of 'TEXT' with verbs of list (5) : 'believe', etc.
I found no such evidence in the Janaske corpus and I had no
chance of checking my intuitions about these verbs with native
speakers. To take just 'believe' and 'belief', it seems to
me that 'In this article, the author believes that—' is
possible, that 'This article believes that—' might be found
acceptable by some speakers, but that 'The paper's belief
is that—' is impossible.

(11) An Experiencer analysis of 'AUTHOR'

In sentences with the verbs or cognate expressions of list
(5) - such as 'believe' or 'belief' - Fillmore would assign
'AUTHOR' the Dative case (1968a, 24), which he now calls
Experiencer (1969a, 116). The first basis for distinguishing
between Agent and Experiencer is conceptual : it is expressed
in Fillmore's definition of these cases. However, I can see
two other reasons for an Experiencer analysis of the subject
of 'believe', etc. One is that, irrespective of details
which I have been unable to establish, the verbs of list (5)
do not seem to accept Locatives as subjects as easily as do
those of list (4). Another difference is that verbs like
'believe' do not accept manner adverbials such as 'briefly'

'concisely', 'carefully', 'verbally', 'in detail', etc. :
'x The author believes briefly that—'.

(111) Arguments for a Topic case.

The following sentences suggest that the notion of
topic must somehow be associated with at least some of the
verbs under discussion. Indeed, some way must be found to
predict the occurrence of the words underlined in the
following sentences :

(11)

(a) covers such areas as ... (J 587)
(b) covers the following topics (J 584)
(c) emphasizing matters ... such as ... (J 715)
(d) the subject of ... is described (J 829)
(e) guides to subject fields such as (J 912)
(f) information on the subject (J 928)

The question is whether a new case is the appropriate
way of capturing such facts. An alternative solution would
be to regard the topic label as a presupposition, which could
be attached to the argument of the verbs in the fashion
proposed by Fillmore 1969a : for instance, 'x discusses y'
for 'y is a topic'. Another solution, of the type suggested
in Langendoen 1970 (78), would be to subcategorize another
case such as the Object case ("the most neutral case") by
means of the topic case : "Object : Topic ". I would object
to both solutions on the following grounds. Such a subcategor-
ization as well as presuppositions always serve to state
restrictions : for instance, the fact that x must be 'human'
in a sentence with ACCUSE. But, with verbs like 'discuss'
what is needed is on the contrary something very neutral and
general, since there is no conceivable limit to be imposed
on the type of things which can be 'described', 'discussed',
etc.

312.

Furthermore, I can think of no better characterization
than a Topic case for the x in the following sentences :

(12)

(a) x is the subject of this paper.
(b) The subject of this paper is x.
(c) This paper is on x.
(d) This paper is about x.
(e) This paper concerns x.

I would agree with George Lakoff in regarding 'be about' and
'concern' as "two-place relations, whose arguments are a
description of a proposition or discourse and the item which
is the topic of that proposition or discourse" (G. Lakoff
1969c, 30) and, within case theory, I see no better way of
capturing this fact than by labelling the x of sentences such
as those of (12) as a Topic case.

If this is correct, the question arises whether such an
analysis can be extended to at least some verbs with
sentential objects : particularly, verbs of list (4) like
'explain', 'say', 'state', 'tell'. An argument in favour of
such an extension is the following : sentences of the type
'He discusses x' (or 'His paper is about x') and of the type
'He says that ...' are BOTH equally appropriate answers
to such questions as 'What is the subject of his paper', or
'What does he talk (write) about ?', or 'What is his paper
about ?'. Furthermore, 'He says something', etc. is an
uninformative answer to such questions. In section 3.2.3., I
shall return to this point and elaborate upon this argument
on the basis of evidence found in the Janaske corpus.

The same extension of the Topic analysis to predicates of list(5) is, I think, to be rejected on the following grounds: in answer to the above questions it is always possible to say, '<u>He says that he believes that—</u>', but hardly, '<u>He believes that he says that—</u>'. In other words, I would say that like '<u>discuss</u>' and other similar verbs, verbs like '<u>say</u>' introduce a Topic, whereas verbs like '<u>believe</u>' are part of the Topic. The same kind of test could serve to establish whether verbs of list (4) are of the same type as '<u>discuss</u>' or not . <u>Question</u> : 'What does he talk about ?' <u>Answer</u> : 'x He says that he discusses—', 'x He says that he states that—', etc. versus 'He says that he warns against—'.

(1v) <u>Prepositions as predicates versus 'essive' analyses.</u>

There is currently much hesitation in TG studies on how to treat prepositions. To take just one example, Langendoen hesitates between adopting Fillmore's <u>essive</u> proposal (for <u>N is N</u> sentences : Fillmore 1968) and the representation of prepositions as predicates, and he mentions sentences such as 'A is in B' = 'B contains A', 'A is about B'= 'A concerns B', 'A is by B = 'B wrote (composed) A' among the "problems and suggestions for further study" (Langendoen 1970, 99). Also consider the following asymmetries

(13)

(a) The text is on x
(a') The text discusses x
(b) x The author is on x
(b') The author discusses x
(c) In the text, the author discusses x
(c') ? In the text, (it) is on (or: concerns) x
(c") ? In this story, it's about x.

A further problem is the existence of '_there is_' and
'_there are_' next to '_be_'. Compare the following :

(14)

(a) This paper is an account of--
(a') _In_ this paper there is an account of--
(b) This book _is in_ three chapters.
(b')This book _contains_ three chapters.
(b'') There are three chapters _in_ this book.

Note that (14) (b'') is different from 'There are many toys
in the box' which was analyzed in Fillmore 1968a (456): it
seems to me that 'three chapters' are part of 'this book'
in a way in which the 'toys' are not part of the 'box'. Note
moreover that '_is on_'does not behave like '_is in_': correspond-
ing to '_is on_' there is no sentence with '_there is/are_' like
(14) (b''). This in fact may be regarded as an argument in
favour of the analysis of '_is on_' as a predicate like '_concern_'.

In the absence of more reliable knowledge about the
questions dealt with in this paragraph, I shall not attempt
to propose an analysis for sentences of the type illustrated
above in (14). I would merely suggest that I see no objection
to analyzing '_an account of—_' and similar nominals in
sentences like (14) (a) and (a') as _N + Topic_ (cp. Stockwell
_et al._'s PS rule 7).

To sum up, I would propose that the verbs under discus-
sion fall into the following broad types; in (c), I write
S for what Stockwell _et al._ 1968 represent as N.

(15)

(a)


```
                          S
         V        Agent        Locative        Topic
      DISCUSS    Author          Text           NP
```

(b)

```
                    S
         V        Locative?        Topic
      is ON         Text            NP
```

(c)

```
                    S
         This paper is    S
                        N      Topic
                  an account of   NP
```

(d)

```
                         S
         V       Agent        Locative      Topic
      STATE     Author          Text         NP
                                              S
```

(16)

(a)

```
                              S
            _____/|_____
           /          /          \             \
          V         Agent       Locative      Object
          |           |            |            NP
          |           |            |            or
        WARN        Author        Text          NP
                                                 |
                                                 S
```

(b)

```
                    S _ _ _ _ _ _
            _____/|\_ _ _ _ _ _ _ - Locative (?)
           /        |    \
          V        Exp.    O
          |         |      NP
          |         |       |
       BELIEVE    Author     S
```
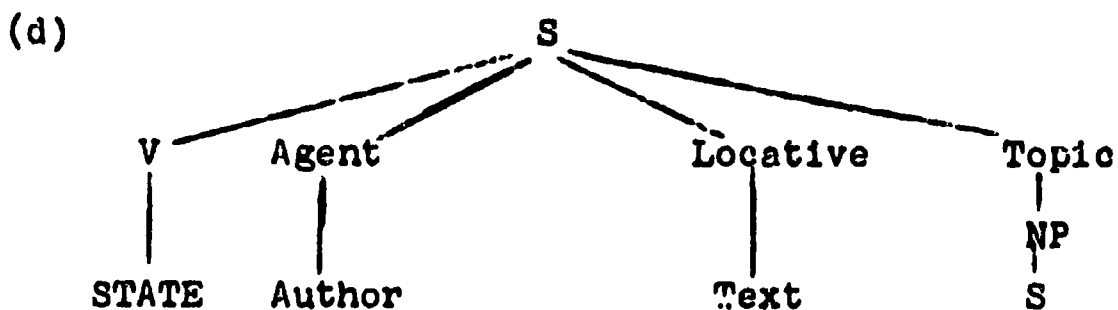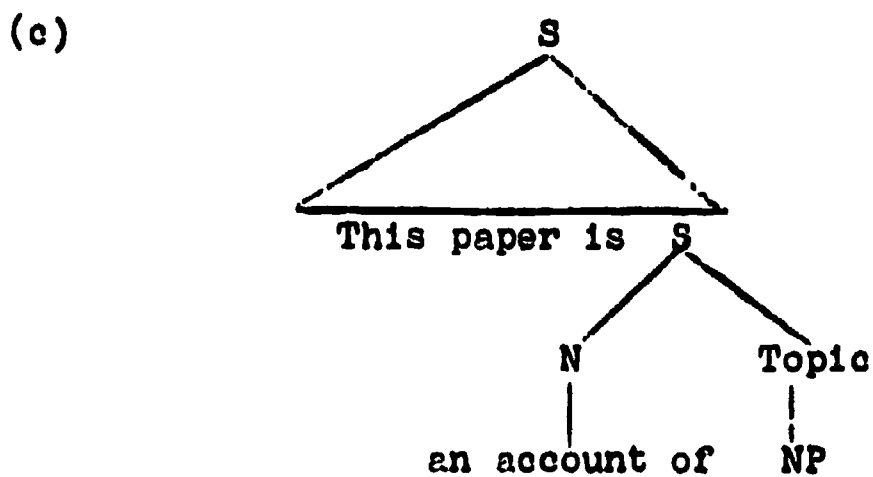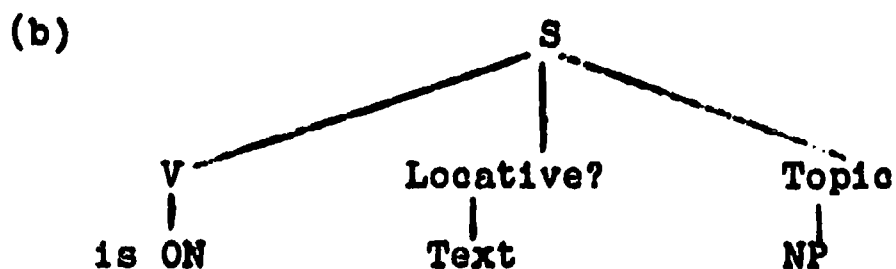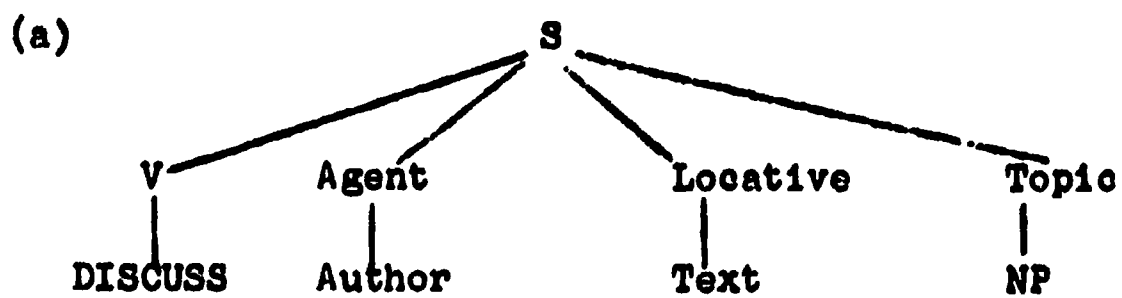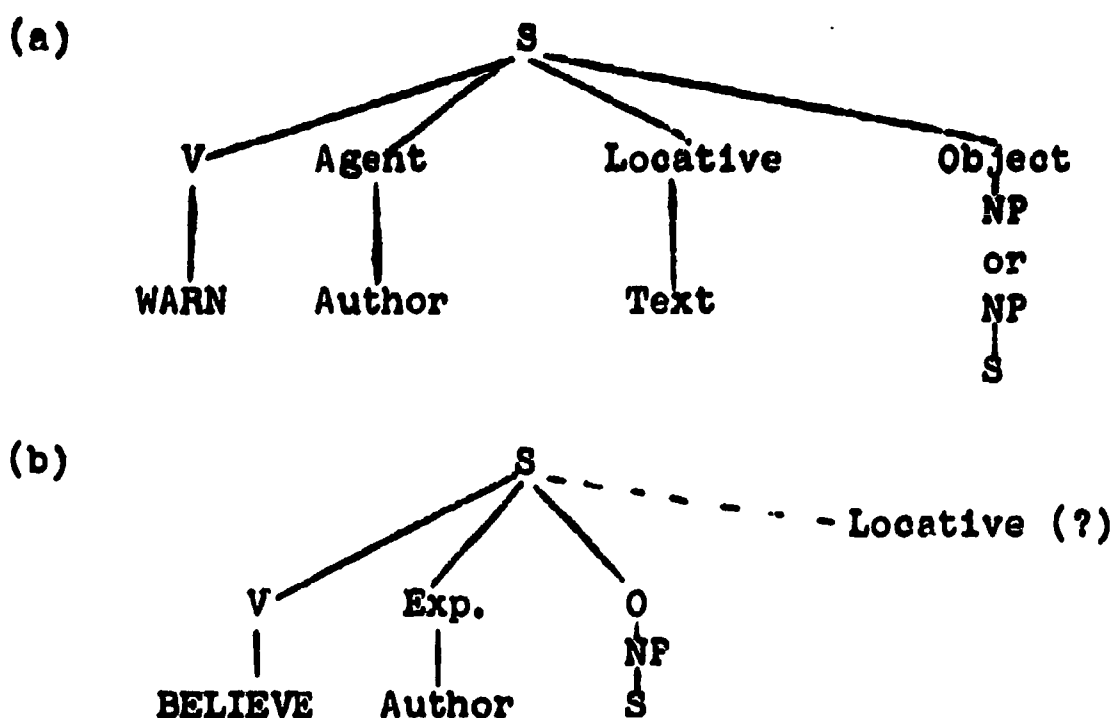
On the basis of such representations, I suggest that rules
could be written that make use of the implication relation
(Lyons and Leech) or of Katz's and others' similar entailment
relation. First of all, such rules would state that a bilateral
implication relation holds between sentences with 'discuss'
(etc.), 'is on' (etc.) and '... is an account of', irrespective
of whether TEXT or AUTHOR is selected as the subject of
'discuss' and similar verbs. Secondly, implication rules could
also be set up to state, for instance, that a sentence such as
'The author explains that a system has be designed...' imply
'The author discusses a system which has been designed...' .
I shall not go into the detailed formulation of such rules,
because it seems to me that they present a serious problem
for case theory. Indeed, I would hypothesize that such rules
would have to say that, for instance, 'x explains that S'
implies that 'x discusses z', where z is what is selected as
subject of S, no matter what case it is. Lastly, if following

Leech we can say that 'John is madly in love with Susan' implies deductively 'John is in love with Susan' (Leech 1969, 36) we can say similarly that sentences like the following imply that, for instance, 'The author discusses x' : 'The paper describes x in some detail', 'The writer describes x in terms of Boolean algebra', etc.

3.2.1.2. Articulate sentences : active.

So far I have dealt with expressions which are fairly simple in their surface form. However, the Janaske corpus contains many expressions, the complexity of which is even more beyond the powers of any theoretical framework known to me than those discussed in 3.2.1.1. And yet, these expressions are intuitively relatable to those of the type 'the author/the paper discusses—'. Three of these expressions are quoted below, in (17) :

(17)

(a) This book fills the need for a text in the way of a general introduction to the subject of — (J 575).

(b) The purpose of this paper is to remove some of the mysteries surrounding— and to show — (J 696).

(c) The inclusion of— represents an attempt to fill at least in part the existing gap (J 281).

Less complex expressions attested in the Janaske corpus can be mentioned here, in the form of the following table (18). Needless to say that not all the combinations generated by this table are attested or even possible. To take just one examples, "Sixteen authors contributed their views on" is attested (J 449), whereas "the paper attempts to contribute a plea for" is not and may perhaps be found unacceptable by some speakers. Still, I have to resort to a tabular presentation in order to avoid having to quote the hundreds of examples on which my tables is based.

Besides I have left out various attested or possible expressions which would need to be characterized by even more specific co-occurrence requirements than those which I have ignored in the following table : for instance, "hold goals/ideas for", lay/place emphasis/stress on", "report a sampling of", "develop the thesis that", "call attention to", "take pains to", "take issue with", "discuss the general picture of", "express belief/confidence that", "reach/point to the conclusion that", "set forth considerations involved in","draw/bring out conclusions", etc.

(18)

SUBJECT POSITION OR RELATABLE POSITION		VERBS OR COGNATE NOMINALS IN OBJECT POSITION OR IN A RELATABLE POSITION.
Subjects	Corresponding Predic tes	
Only 'AUTHOR' as subject	appends elaborates with (?)	account of analys- appeal-
Only 'TEXT' or 'THIS' as subject	is followed by is prefaced by	compar- defin-
	is addressed to (Ving) is an attempt to is designed to is devoted to is an effort to is intended to/as	demonstrat- descri-(ption/bing) determ- discern- discuss- evaluat-
	consists of contains covers (?) is included	examin- exampl- expla-(nation/ining) explor-
'AUTHOR' or 'TEXT' (or 'THIS') as subject or as genitive of nominals	begins with concludes with ends with (cp. terminate text)	facts about idea(s) illustrat- list-
	aims at aim is to attempts to goal is to purpose is to seeks to tries to includes	notes on mention of observ- outlin- a paradigm for a plea for (the following) points : report-

±	contribut- furnish- giv- mak- offer- present- provid- show-	review- sampl- show- stat- study- summar- supplement- survey- views on

It seems to me that within the limits of current theories - particularly case theory - the following points can be made here in connection with the above table. I shall ignore various details, some of which are rather mysterious : compare, for instance, "The text is prefaced by a discussion of___", and "?x The text is appended by a discussion of___".

First, the same grammatical alternatives as those noted above in connection with 'discuss', 'believe', 'is on', etc., exist for the superficially more complex expressions just tabulated - namely just AUTHOR or just TEXT or both as possible subjects. Besides there is a certain amount of semantic parallelism cutting across the various grammatical alternatives : 'succession in text' (appends, is followed by, begins with), 'purpose' (is addressed to, aims at), 'inclusion' (contains, includes). Second, it seems reasonable to posit that expressions like 'make an attempt/an effort' and 'contribute/ furnish/give/present... an analysis/etc.' have the same case representation which is of the form V (Agent, Goal) - where attempt, analysis, etc. are exponents of the Goal case. Third, it also seems reasonable to extend the Agentive analysis of 'author' in 'The author discusses___' to all the expressions which require such an extension : 'The author begins with a discussion of___', 'The author attempts to discuss___', 'The author presents a discussion of___', and 'The author attempts to present a discussion of___'. Similarly,

the locative analysis of 'paper' in 'The paper discusses___'
and in 'In this paper, the author discusses___' could be
extended to the following types of expressions :

(19)

(a) In this paper, the author begins with a discussion of—
(a') This paper begins with a discussion of—
(b) In this paper, the author attempts to discuss—
(b') This paper attempts to discuss—
(c) In this paper, the author presents a discussion of—
(c') This paper presents a discussion of—
(d) In this paper, the author attempts to present a
discussion of___
(d') This paper attempts to present a discussion of—

Sschematically then, I would tentatively propose the
following analyses :

(20)

(a) Only AUTHOR as subject.

(b) Only TEXT as subject

S
V — CONSIST of / INCLUDE
Locative — Loc — Text
SO — N — ACCOUNT of
Topic

S
V — is an ATTEMPT to
Locative — Text
Goal — SO — V — DISCUSS
Topic

(c) <u>AUTHOR or TEXT as subject</u>

S
V — INCLUDE
Agent — Author
Locative — Loc — Text
Loc — SO — N — discussion
Topic

S
V — BEGINS with
Agent — Author
Locative — Loc — Text
Loc — SO — N — discussion
Topic
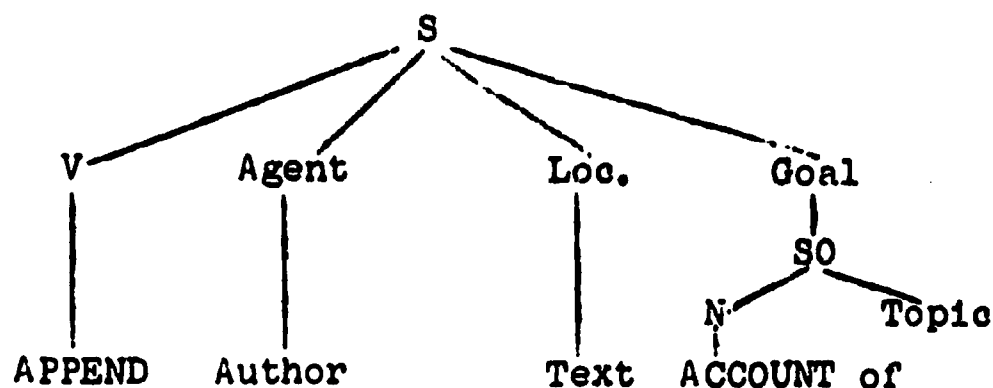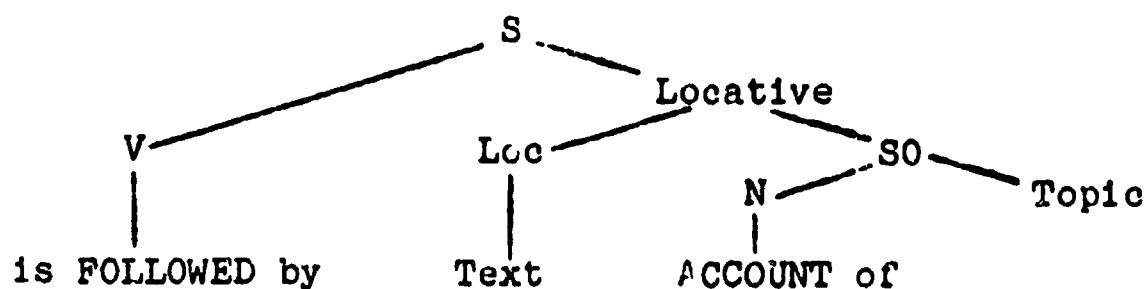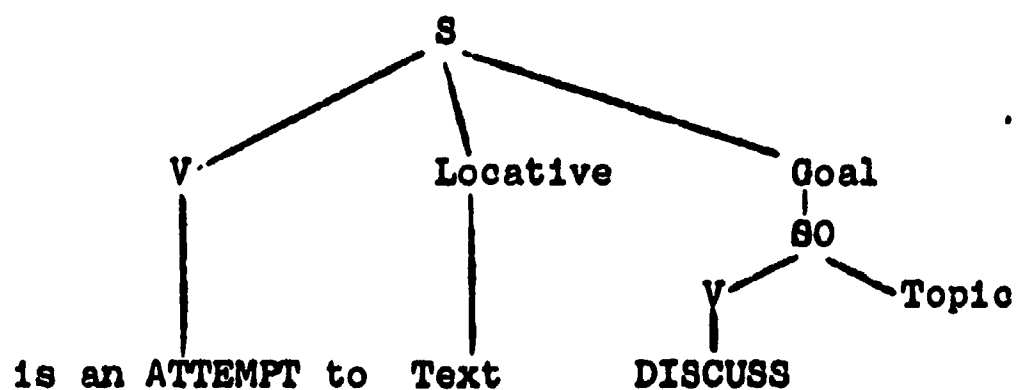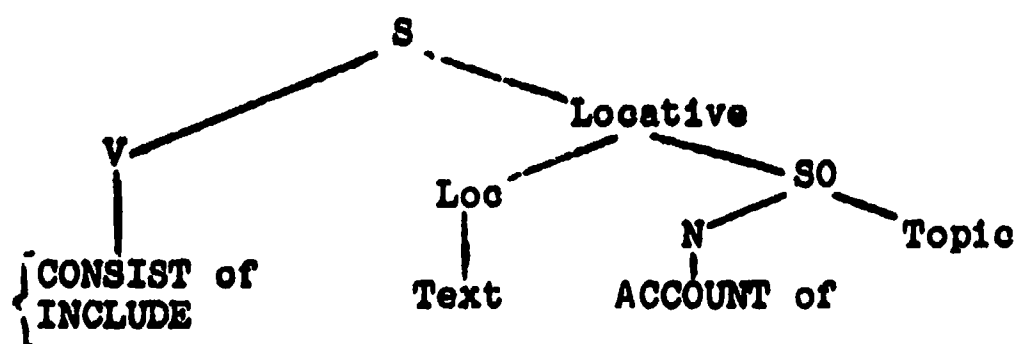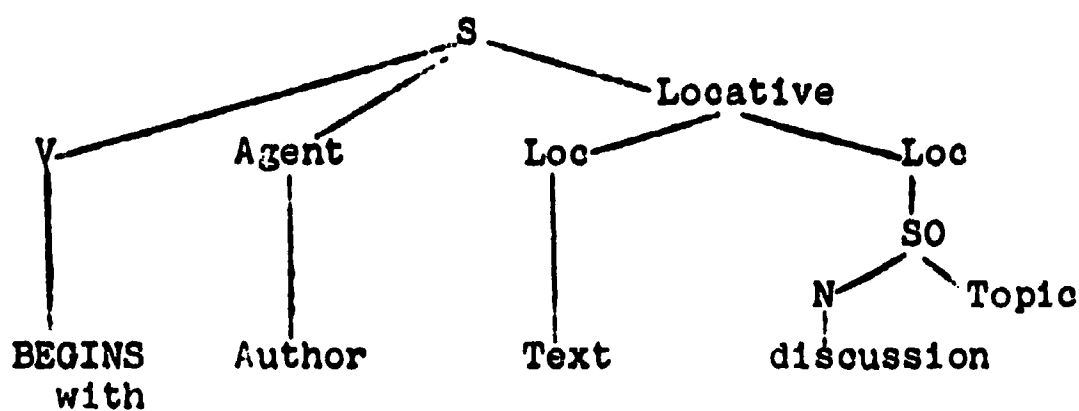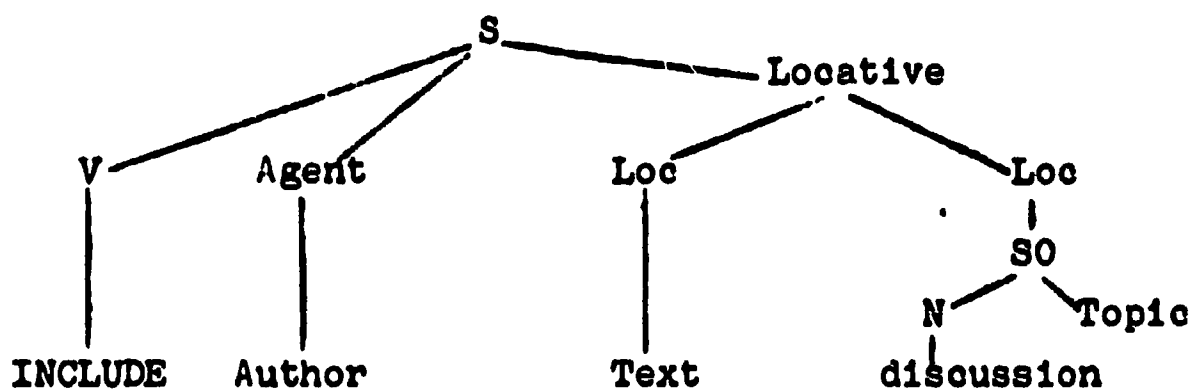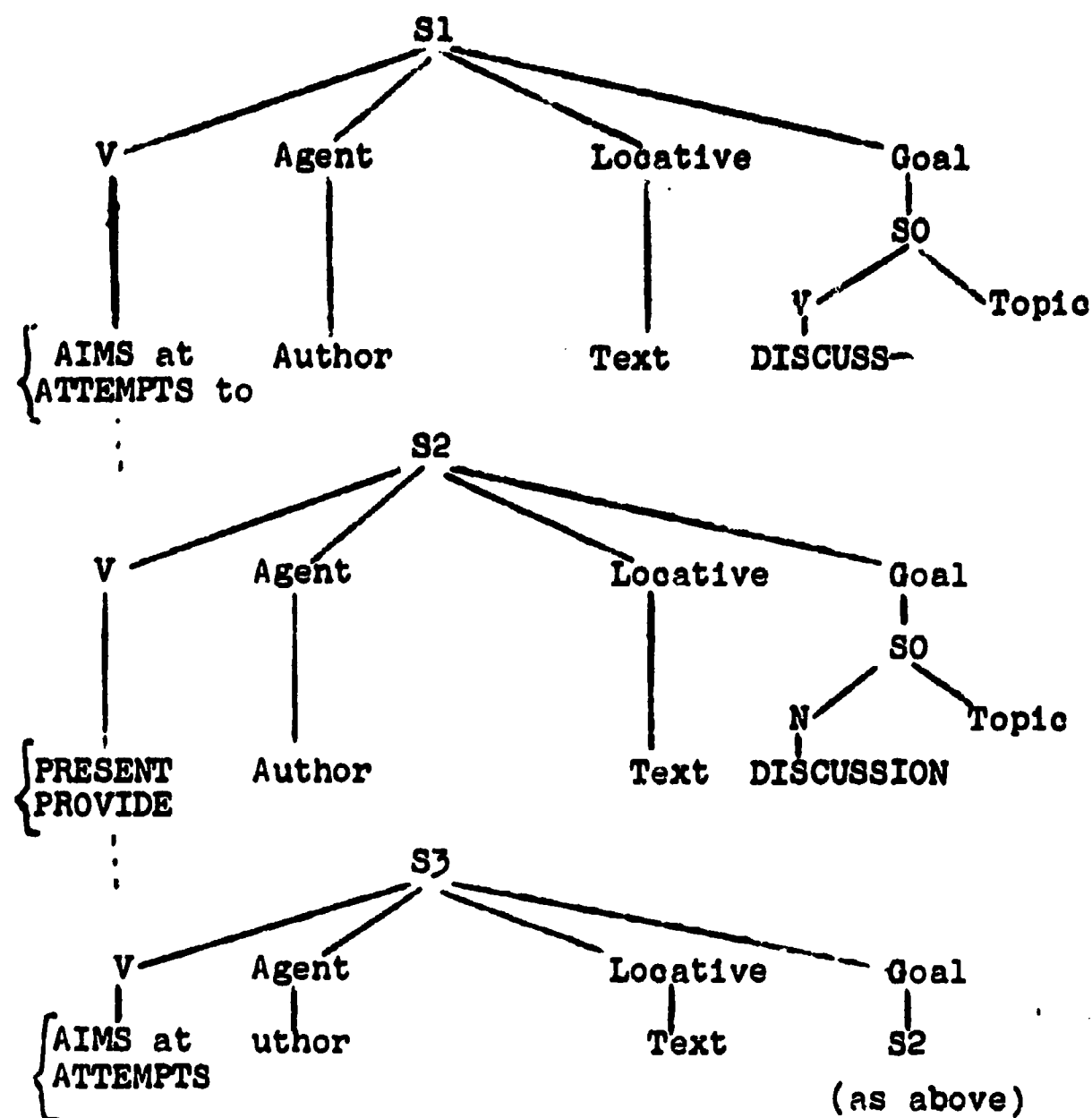
Here again, I suggest that implication rules could state, for instance, that S3 implies S1, S1 implies S3, and presumably that S1 and S3 imply S2. Similarly, all the sentences diagrammed in (20) can be said to imply SO, that is, "DISCUSS + Topic", or the like, but not vice versa.

I wish to end this section by making two further points. The first one concerns the transformation known as equi-NP deletion by which a sentence of the type 'John tries to read' is derived from a deep structure of the form 'John tries (John read)'. I have omitted various problematic details in the diagrams given in (20). Consider, for instance, a sentence of the type

I have called S3 : 'In this paper, the author attempts to
provide a discussion of—'. Suppose, furthermore, that we
wish to say that 'account' or 'discussion' in this sentence
is actually a verb in deep structure. Under such an assumption
S3 would have to be diagrammed as in (21), since the three
verbs of this sentence accept both Author or Text as subject:

(21)
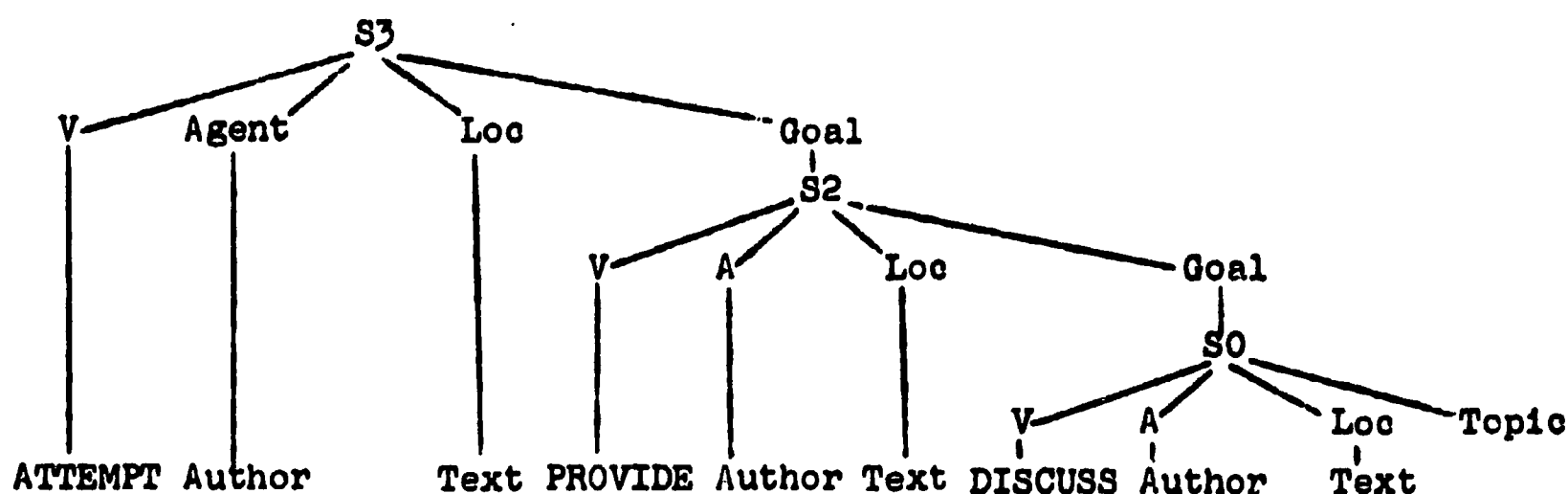
But note that such a representation would require four
noun-phrases eligible as subject to be deleted, namely the
two Authors and Texts associated with DISCUSS and PROVIDE
respectively. I am not acquainted with any solution proposed
for this problem in the literature. Whatever the solution may
be, I shall return to some aspects of this problem in 3.2.3. :
in particular, I shall argue that it is difficult to represent
semi-articulate sentences such as 'Discusses—' and 'A discus-
sion of—' (standing at the beginning of an abstract) as I
have represented SO in the above diagram. In other words, I
shall argue against deriving these semi-articulate sentences
from corresponding articulate sentences such as 'In this paper
the author discusses—'.

My second point is perhaps relatable to the problem of
ATTEMPT (to) in the above diagram. It has to do with various
purposive sentences. Compare the purposive sentences of the
type tabulated above in (18) (for instance, "The specific
aim of the study is to discuss___" : J 1051) with those listed
in the following table :

(22)

'DISCUSSION OF	S1	PURPOSE	S2
- Report of	a(premilinary) study	to	obtain some empirical data on— (J 842)
- Report of	a study by(...)	for the purpose of	determining— (J 852)
- ∅	A study of(...)	for the main purpose of	analyzing— (J 879)
- A report of	a three weeks' trip.	The three aims of the trip were	1)to examine__ 2)to study— 3)to control— (J 887)
- ∅	This article is WRITTEN	to	help— (J 916)
- Report of	a study CONDUCTED at__	for the purpose of	examining— (J 919)
- ∅	∅ DESIGNED	to	acquaint__ (J 920)
-	A chart is GIVEN	to	illustrate— (J 983)
-	—were examined	to	determine— (J 938)
-	Explores__	with a view to	discovering— (J 522)

I shall leave the problems raised by·the first column un-
discussed in the present. thesis Now, if we consider column S1 in
the table, the nouns in this column are relatable to sentences.
which seem.'to   be of the type Verb + Goal : "make a trip",
"write an article", "design/conduct a study", "give a chart".
Others could be analyzed as Verb + Topic : "examine—",
"explore—". Such data, it seems to me, raise the following
question. Consider sentences of the type 'the author's purpose
is to discuss—' (or 'The purpose of the paper is to discuss—'),
and, say, 'This paper/The author attempts to discuss—'. The
question which arises is whether the deep structure of such
sentences should not be a two-clause sentence roughly of the
form'S1 for the purpose of S2'.

I can think of two arguments in favour of such an alterna-
tive analysis. First, while it is possible to say "The study
analyzes—", "The purpose of the trip/study is to analyze—",
it is not possible to say "x The trip analyzes—". On the
other hand, purposive expressions impose various other
restrictions on the predicates depending on them : e.g.
'x The author's objective is to feel that—'. Second, nouns
like 'paper', 'author', 'trip' and the like are by no means
the only grammatical category that can enter into the frame
'The purpose of—is to' : for instance, "The essential purpose
of literature searching is to locate..." (J 773). In chapter
3.4., I shall propose a (two-clause analysis for such sentences
which may also be applicable to some of the examples given above
in (22).

Applying this two-clause analysis to sentences like 'The purpose of the paper/author is to—' would of course require that 'paper' and 'author' in such sentences would be represented in deep structure as sentences or as cognate nominals ie.g.('the) writing this paper' . While I have r idea of how this could be done, I feel that examples like the following make such a proposal plausible : compare "At the time of this report" (J 1) and "At the time this report was written" (J 121).   The difficulty with positing deep structure sentences for surface nouns like 'paper' is that there is no known way of deciding which (set of) "appropriate" verbs would have to be posited in such sentences : 'write', 'prepare', 'publish' (this paper), etc. (see F rris 1968, 164 f. on "appropriate words").

3.2.2. Articulate sentences : passive.

I only wish to mention briefly the following points :

(1) There are various expressions which I believe to be superficially as well as semantically too complex to be tractable and I shall not attempt to analyze them. One example is : "In addition to references dealing with— , entries have been included covering the subject of—"(J 1000).

(11) There are a number of expressions of which it is difficult to say with certainty whether they are passive. Some examples are listed in (23). Both (23) (a) and (b) have been treated as actives in the preceding section. If the active counterpart is regarded as unacceptable ('x A brief account prefaces—', (a) cannot be regarded as a passive. Similarly, (b) sounds more natural to me than its active counterpart: ? 'A outline for — follows a discussion of—' ; I would say that the intransitive active use of 'follows' is more acceptable, as in : 'An outline for — follows, emphasizing —'. I have no explanation for such facts. Similarly, the active counterpart of (c) sounds a bit funny to me : 'In these pages the author compresses/has compressed a general review of—'.

<div style="text-align:center">(23)</div>

(a) — is prefaced by a brief account of— (J 122)
(b) A discussion of — is followed by an outline for— (J 136).
(c) In these pages is compressed a general review of— (J 1007).
(d) ... the authors have included—.— are included also. (J 975).
(e) Also included are tables which indicate— (J 654).
(f) Included in the paper are : a detailed discussion of— and mention of— (J 116).
(g) In it is outlined— (J 891) .

A similar problem about deciding whether a sentence is active or passive arises with expressions such as 'is designed to', 'is devoted to', 'is intended to/as' which were treated as actives in the preceding section. As Fillmore once put it, "Stative adjectives can be derived from (...) change of state verbs" such as break : 'The window is broken' can be understood either as a passive or as a description of a state. Similar sentences with other verbs, such as 'hit' ('The window was hit'), can only be understood as passives. With reference to 'is designed to', etc., I would say that such sentences are ambiguous like the sentence with 'break' but I can think of no explanation for this fact (Fillmore 1967, 18-19).

(iii) Next to those exemplified in (23), I have found various constructions which I shall not attempt to analyze. Compare the constructions superficially similar to passives in (24) and (25) with those given in (26) and (27) respectively.

(24)

(a) Listed and annotated are— (J 826).
(b) Announced in this paper are— (J 985).
(c) Appendixed are— (J 1004).
(d) Also reported is— (J 1115).

(25)

(a) Arising from this study is— (J 919).
(b) x Mentioned is that—
(c) Mention is made of the fact that—
(d) The fact is mentioned that—

(26)

(a) Finally, there is presented an explanation of—(J 421).
(b) There is thus achieved the general description of—
(J 963).

(27)

(a) There is a very useful supplement of— (J 38).
(b) With regard to—, there was a plea for— (J 525).
(c) There is a general examination of— and a detailed description of— (J 664).
(d) However, there are just as many warnings on— (J 923).
(e) There is a discussion of— (J 993).

(f) <u>There is</u> also a general estimate of—(J 1020).

(g) <u>There are</u> numerous illustrations and a two-page bibliography (J 1063).

(h) <u>There is</u> agreement with the general belief that— (J 1069).

(iv) According to Langendoen (1970, 142) "The girl (whom my cousin married) became pregnant" is derived from "My cousin married a girl. She became pregnant". Thus, a sentence such as (28) (a) would be derived from a deep structure having the form illustrated in (28) (b) - namely "a discourse consisting of two sentences"：

(28)

(a) The system described employs ... (J 35).

(b) A system is described. It employs—

Langendoen (<u>loc. cit.</u>) claims that in the sentences quoted in (29) (a) and (b) "the temporal order" of the sentences "suggests a corresponding temporal order to the events described" : (a) "describes a situation in which a girl becomes pregnant after marriage", and (b) "a situation in which she becomes pregnant before marriage". Besides, he claims that "this semantic difference between the discourses carries over into the sentences with relative clauses" — (29) (c) and (d) respectively :

(29)

(a) My cousin married a girl. She became pregnant.

(b) A girl became pregnant. My cousin married her.

(c) The girl whom my cousin married became pregnant.

(d) My cousin married a girl who became pregnant.

Irrespective of what one may think about Langendoen's linguistic intuitions here, the fact is that no temporal order is present either in the sentences with relative clauses or in the corresponding sentences given in (30).

When such an order is present, as may be the case in some examples of (31), it seems to be entirely due to the meaning of the sentences or clauses involved, and not to the grammatical construction.

(30)

(a) A system which employs (...) is described.
(b) The system which is described employs—.
(c) The system described employs— (J 35).
(d) A system is described. The system (It) employs—.
(e) The system employs— This system is described here.

(31)

(a) The classification scheme <u>presented here</u> had its origin in... (J 33).
(b) Documentary activity will be coordinated at national level by the <u>proposed</u> Institut für Dokumentationswesen (J 47).
(c) Arising from the study is a suggested standard for abstract journals <u>which is also reported in this paper</u> (J 919).
(d) — are reported in this paper (J 1028).

(v) I shall assume that what has been said in the preceding sections about <u>implication</u> rules can hold for passive constructions, provided the suitable case analyses are available. Thus, a presumably bilateral implication relation can be said to hold between the following sentences:

(32)

(a) Efforts are devoted to exploring <u>x</u> (J 120).
(a') x is/are explored.
(b) An attempt is made to give a general picture of x (J 590).
(b') x is depicted/described.

(vi) Most of the undoubtedly passive constructions in the
Janaske corpus are of the form '——is/are mentioned' or of
a slightly more complex form such as 'Mention is made of——'.
The forms which I have found will be listed below. What is
important to note here is that the great majority of these
sentences contain no mention of 'AUTHOR' or 'TEXT'. I have
found only one sentence in which both are mentioned (33a).
One clear instance of a passive sentence containing a
presumably Instrumental expression has been found (33 b),
and only about twenty sentences with a locative preposi-
tional phrase at the beginning (as in 33 c), or at the end
(as in 33 d), or an adverb (as in 33 e) :

(33)

(a) — is explained by G. Meyer in an article entitied—
    (J 798).
(b) — are listed and assessed in the light of— (J 599).
(c) In this paper, — are described (J 680).
(d) — is dealt with in this article (J 789).
(e) — are reported herein (J 298).
(e) — is here presented (J 486).

The verbs and verbal expressions which I have found in
passive constructions in the Janaske corpus are listed below
in (34) and (35). In the latter table I omit the statement
of various co-occurrence restrictions which would specify,
for instance, that 'emphasis' is 'given to', 'put on',
'placed on', that 'questions' are 'raised', 'asked',
'answered', that 'note' is 'made', that 'comparisons' are
'made' or 'drawn' (J 1077), that 'a belief' is 'expressed',
etc.

(34) IS/ARE :

advanced
advocated
analyzed
appended
appraised
arranged
assessed
attributed to
believed(to be)
claimed
compared to
computed
considered
covered
criticized (as)
dealt with
decried
defined as
delineated
depicted
described
detailed
determined
diagrammed
discussed
displayed
distinguished from
elaborated
emphasized
enumerated
enunciated
estimated
evaluated
examined
expected to
explained
found
given

groped for
grouped
identified
illustrated
included
indicated
investigated
justified
listed
mentioned
noted
outlined
pointed out
presented
presumed
proposed
purported to
quoted
recommended
regarded
rejected
reported
reviewed
shown (to)
sketched
stated
stressed
studied
suggested
summarized
tabulated
touched upon
traced
treated

(35)

	(a) +		(b) IS/ARE :
account		added	
analysis		advanced	
annotations		appended	
appeal		drawn	
attempt		given	
a catalogue		included	
charts		made	
comments		offered	
comparison		presented	
conclusions		provided	
considerations		outlined	
data		submitted	
details		summarized	
description		supplied	
distinction			

account
analysis
annotations
appeal
attempt
a catalogue
charts
comments
comparison
conclusions
considerations
data
details
description
distinction
dream
examples
estimate
evaluation
explanation
generalizations
glossary
guide lines
illustrations
indication(s)
indexes
information
list
observation
plans
proposals
recommandations
(general) remarks
report
resume
review
samples
statistics
suggestions
survey
synopsis
table(s)

added
advanced
appended
drawn
given
included
made
offered
presented
provided
outlined
submitted
summarized
supplied

I suggest that the subject of passive sentences with at least
some of the verbs listed in (34) may be analyzed as the Topic.
For the expressions listed in (35) what I propose to describe
as a possible Topic has the grammatical form of a prepositional
phrase depending on the subject head noun : for instance, x
in 'an account of x is given',or in'an account is given of x'

In the rest of this section, I shall be concerned with
a problem which can be stated in the form of the following
paradox. As will be remembered, 'AUTHOR' and 'TEXT' are mostly
not mentioned in the passive constructions under discussion;
on the other hand, it seems that 'AUTHOR' and 'TEXT' are
conceptually present under the influence of the context, but
that it would be a mistake to posit them in deep structure.
This suggests that passive constructions such as 'x is
discussed' would be relatable not to the corresponding active
sentences ('In this paper, the author discusses x') but rather
to semi-articulate sentences of the type dealt with in the
next section (3.2.3.) : 'A discussion of x', and 'Discusses x'.
They would have a deep structure of the form Predicate +
Topic. Now, Fillmore suggests that sentences like 'The arrow
hit' are instances of "definite" pronominalization. He argues
that 'The arrow hit' and 'The arrow hit IT' are paraphrases
and that in both, "the speaker expects the identity of the
'target' (Goal) to be already known by the addressee". More-
over, he proposes that such facts should be represented in the
lexical entries, as idiosyncratic properties of predicates
such as 'hit'. The following is, I think, a strong argument
against such a conception as far as the passive sentences
under discussion are concerned. When 'AUTHOR' and 'TEXT'
are not mentioned, you cannot refer back to them by means
of IT or HE (SHE) :

(36)

(a) x x is described. <u>It</u> also describes x.
(b) x x is described. <u>He</u> (<u>She</u>) also describes x.

Moreover, such a fact cannot be regarded as an idiosyncratic property of the predicates in question, considering that 'AUTHOR' or 'TEXT' are so to speak always omitted from the passive sentences under discussion, and that this is the case with all the predicates attested. This suggests that conceptually incomplete structures of the type <u>Predicate + Topic</u> should be generated by the base of the theory (as should be Jespersen's "<u>Thank you</u>"), and that our ability to relate the corresponding sentences to their 'complete' active counter- parts should be accounted for not by transformations, but by a device of an unknown nature : presumably one of the type envisaged by George Lakoff (1971), which would PAIR representa- tions of <u>contexts</u> with representations of <u>sentences</u>. Such a device would give recognition to Jespersen's observation that

> Only bores want to express everything, but even bores find it impossible to express everything (...) in the most everyday remarks we suppress a great many things (...) "Two third Brighton return" stands for something like : "Would you please sell me two third-class tickets from London to Brighton and back again, and I will pay you the usual fare for such tickets". (Jespersen 1965, 310).

The usual absence of an Agent in the passive sentences under consideration may result in ambiguities which can often not be resolved linguistically. Moreover, the same kind of am- biguity arises with nominals (Jespersen's nexus-substantives). This happens when it is not clear whether the understood Agent of a predicate is what Jespersen calls "the generic person" or some definite person, for instance, the author of the abstracted publication (Jespersen 1966, 317). Thus, expressions such as '<u>It is assumed that—</u>' and '<u>the assumption that—</u>' can have two readings - and occasionally three -

in the passages quoted below. These readings can be paraphrased respectively as 'The author assumes that—', and 'It is generally/usually assumed that—' or 'The abstractor assumes—'. To show how this ambiguity arises, it is necessary to quote more than one sentence. Compare the passives in (37) (a) and (c) with the nominals in (b) and (d); in both (a) and (b), it is impossible to know for certain who "assumes". In (c), the abstractor, the author and the "generic person" all seem to be eligible as implicit Experiencer of 'believe'. In (d), we may regard "the Library Association in London" as an additional possibility, presumably ruling out the abstractor (because of 'The belief').

(37)

(a) (First sentence of the abstract; 'Author' is the implicit Agent of 'outline':) Plans are outlined for organization of institutes for medical statistics and documentation (...). It is assumed that all clinical hospital departments (....) will initiate... (J 545).

(b) (First sentence of the abstract; 'Author' is the implicit Agent of 'discussion', but not necessarily of 'assumption':) A discussion of computers based on the assumption that they are, could, or should be considered the ultimate in information processing devices (J 619).

(c) This is a dictionary of the special meanings given to over 400 words and phrases as used by documentalists. Designed to assist those just entering the field, it is believed that it will also give some stability to the word usage by documentalists already in the field. (J 1058).

(d) Approach to the training of scientific information specialists as seen by the Library Association in London. The belief is that librarian and information officer use basically the same materials and techniques... (J 747).

Other examples of ambiguous passives are given in (38).
In all these examples, the first sentence of the abstracts is
quoted and is unambiguous in that 'AUTHOR' is the Agent
understood from the context. For reasons which I am unable
to explain, some sentences that _follow_ the sentence with the
ambiguous passive may be _unambiguous_, as seems to be the case
in (b) and (c). Let us say that what I mean by 'context' here
is the format adopted in the Janaske bibliography to present
the abstracts : the name of the author, followed by the title
of the abstracted publication, itself followed by the abstract
(Appendix 4).

(38)

(a) A cost comparison is made between manual and mechanical
selection operations in a small company (...) Costs
for mechanical selection _are found to_ be double those
for manual methods, but _are justified_ on the basis
of ... (J 382).

(b) Discusses potentialities of a retrieval system (...)
Capacity of system in present form _is estimated_ at
(...) Details of system _are not described._ (last
sentence : J 238).

(c) The Delta punched card is of the Batten-Cordonier
type. _It is recommended_ in cases where ... (J 594).

(d) The first Soviet tests of machine translation were
conducted in 1955 (...) _It is expected that_ by 1961...
(J 624).

(e) Reports on a technique for literature indexing (...)
This "probabilistic indexing" method _is said to_
allow (...) Finally, the results of a small-scale
experiment _are given and analyzed_ (J 640).

(f) This is a third interim report on (...) Because
it is possible to describe one concept by utilizing
several different verbal expressions, _it was found_
necessary to develop one metalinguistic 'common
denominator ... (J 707).

When there exist adverbs   corresponding to the verbs
under investigation, it seems that they are potentially
ambiguous like the verbs in the passive. Thus, '_admittedly_'
can be ambiguous between 'the author admits that—', 'It is

generally admitted that—' and perhaps 'the abstractor
admits that—' in (39) (a); something similar seems to happen
with 'consequently' in (b) because there is not clue as to
who draws the consequence, and in (c) with 'deplorably' -
perhaps because of special properties of believe (Langendoen
1970 on opacity : 114); in (d) and (e), however, there is
no doubt as to the author been conceptually involved in
'critically' and in 'emphatically' for reasons that are unclear
to me.

<div align="center">(39)</div>

(a) States that when (...) it is time for librarians
and documentalists to stop questioning the value of
thesis literature and its control. The practices
with respect to this body of scientific literature
(...) are discussed. Admittedly, progress has occurred
through... (J 92).

(b) The financial success of documentation activities
is difficult to assess. Consequently, it is useful to
calculate the costs... (J 57).

(c) The author believes that the standards are deplorably
low (J 293).

(d) Discusses the question of who should prepare abstracts
and indexes and examines critically ... (J 399).

(e) Fundamental problems of (...) are discussed (...)
It is stated emphatically that ... (J 47).

3.2.3. Semi-articulate and inarticulate sentences.

This section is concerned with semi-articulate sentences
of the type 'Discusses—' and 'A discussion of—' (I ignore
the type 'Presents a discussion of—', which also exist) and
with inarticulate sentences, that is, sentences consisting
of just a noun-phrase ('Three representative problems of...'
J 4), and, more broadly, sentences which are not introduced
by 'Discusses—' and the like. I shall just mention briefly
the following points :

(1) Constructions which have no status - or an unclear status
- in case theory :

(40)

1. **Instrumental expressions** (?)

   (a) A discussion, from actual practice, of— (J 385).
   (b) A detailed discussion (using specific examples)
       of— (J 937).
   (c) A discussion of— based on the assumption that—
       (J 619).

2. **Point of view**

   (a) Lists— by province or state... (J 30).
   (b) Discusses — from the point of view of ... (J 1046).
   (c) —a view from the reference desk. A view of—
       from the standpoint of— (J 312).
   (d) A comparison of— with respect to— (J 630).

3. **In the form of**

   (a) ... communications in the form of comments... (J 987)
   (b) Cp. : Data are presented in summary form (J 945).

4. **Goal** (?)

   (a) An introduction for those new to... (J 717).
   (b) Cp. The report written for laymen as well as for
       scientific specialists... (J 992).
   (c) Cp. Directed to the librarian, this article ...
       (J 1059).
   (d) A plea to all developing an information retrieval
       system (J 663).

5. __Experiencer__ (?)

   __The author__'s concept of... (J 242).

6. __Source__ (Fillmore 1969a)

   ... comments __from individuals__ concerning... (J 987).

7. __Agent__ (cp. Stockwell __et. al.__ 1968)

   (a) A statement __by Senator H. Humphrey__ ... (J 1009).
   (h) Confessions __of   Cleveland documentalist...__
       (J 497).

(11) Some verbal forms have no cognate nominal with the same
     meaning and vice versa. In other cases, the grammatical
     constructions required by the cognate forms are different.
     Besides, some cognate forms can be unacceptable as such in
     __semi-articulate__ sentences :

<div align="center">(41)</div>

	Verbs	Nominals
Fully Equivalent	Lists Reports on	(A) List of (A) Report on
Semantically equivalent	∅ Advocates— Concerned with Deals with ∅ ∅ Gives ∅ Shows Tells of/how Traces (the history) Urges	(An) Account of ∅ x Concern with ∅ (A) directory of An essay on x Gift (A) plea for x Show ∅ x Traces of x Urges
grammatically not equivalent	Explains that Notes that	An explanation of A note on
Difference in semi-articulate sentences	Concludes that Conceives of Observes that Points out that Remarks that	x(A)conclusion that (cp.:the- is that) xThe concept of (cp. The author's concept of) xAn observation that (cp.The-is made that) xA point that(cp.The- is made that) xA remark that(cp.The- is made that).

The following nouns and verbs that function as main
predicates in the semi-articulate sentences under investi-
gation are given in (42) and (43). Contrary to what might
be expected, those that were found in titles are the
minority.

<div align="center">(42)</div>

An account of
An analysis of
An appeal for
Bibliography on/at
A Collection of
Comments on
A comparison of
A compilation of
The author's concept of
A consideration of
Considerations/on/for/of/
        in/relating to
Confessions of
Contents :
Critique on
Description(s) of
Dictionary of
Directory of
A discussion of
A discourse a la ...on
A dissertation on
The distinction between
Documentation on
—encyclopedia
An essay on
An estimate of
An explanation of
Facts and figures on

Guide through/to
Guide lines for
Glossary of
Introduction to
Index tc/of
An investigation of/into/for
List of
A listing of
A manual of
A memorandum
Material on
Notes on
Observations on
Outline of
Papers on
Plea for
A presentation of
A projection of
Proposals of
A proposal for/to
Recommendations for
Refutation
Remarks on
Report of/on
Review of
Research on
Sketch of
Study of/on
Suggestions for
Summary of/on
Survey of
Some thoughts on
A view of

(43)

Advocates	On
Analyzes	Outlines
Annotates	Points out
Assumes	Predicts
Chronicles	Presents
Comments that	Proposes
Concerned with (the idea that)	Provides
Concerns	Recommends
Contends that	Regards
Concludes that	Reports on
Contains (chapters treating)	Reviews
Covers among others such	Says
topics as	Shows
Deals with	States
Defines	Stresses
Describes	Suggests
Discusses(matters of)	Summarizes
Envisions	Surveys
Examines	Tells of/how
Explains	Urges
Explores	Views
Feels that	
Gives	
In(four sections):	
Introduces	
Lists	

In order to understand the purposes of the present discussion, it is necessary to remember that under the standard methodology inherited from TC approaches to syntax, surface sentences are usually derived from their <u>most complete or explicit</u> representations. In other words, deep structures generally represent various entities which can be left un-expressed or implicit in the surface realizations of sentences. Jespersen – who, incidentally, does not "recognize as sentences" such expressions as "book-titles", on the ground that they "occur in writing only and thus fall outside language proper" - grants that "spoken language may indulge in many suppressions" (Jespersen 1965, 311), and points out the difficulty of reconstructing some semi-articulate and inarticu-late sentences : "an old-fashioned grammarian", he says,

will be inclined to explain incomplete sentences "by his panacea, ellipsis". But as Jespersen notes, it is sometimes difficult to do this : "what is understood in (...) "special edition !" ? "I offer you..." or "Will you buy... ?" or "This is... "?" (op. cit., 307).

In the rest of this section, I wish to examine to what extent it is possible and justifiable to reconstruct the semi-articulate and inarticulate sentences under discussion. I shall first consider the semi-articulate type, and assume that what can be said about them is also true of the passive constructions in which 'AUTHOR' and 'TEXT' are not mentioned. Indeed, it seems to me that the expressions underlined in (44) are equivalent. Moreover, the question arises whether the titles - which I regard as inarticulate sentences - are rot equivalent to the semi-articulate sentences that follow (the first sentence of the abstract) :

(44)

(a) New kinds of indexes (title)
    Discusses new chemical indexes which ... (J 83).
(b) New role of machines in document retrieval...(title)
    A discussion of the role of machines in ... (J 100).
(c) The case for centralized internal communications (title)
    Discussion of the need for ... (J 125).
(d) Efficient use of the IBM file of ASTM powder x-ray
    diffraction data (title)
    The use of the IBM file of powder x-ray diffraction
    data ... is discussed.

The technical notion of TG linguistics which is relevant here is that of (nor)recoverable deletion (Katz and Postal 1964; Chomsky 1965, 144-5, 177-82). Formulated in terms of this notion, Jespersen's problem is whether or not it is possible to analyze (i.e. reconstruct) the sentences under discussion in such a way that the reconstructions meet specified conditions allowing for the derivation of these sentences by deletion rules of the so-called recoverable type.

The following passage from Levin 1971 can serve to sum-
marize the question of recoverable deletion; note, in
particular, the statement of three types of conditions
under which a given deep-structure representation lends
itself to deletion transformations :

> In Katz and Postal (1964, pp. 79-81) the principle was
> stated that only those deletion transformations should
> be permitted in a grammar which result in unique
> distortions of a Phrase Marker - in other words, only
> such deletions as are recoverable given the resulting
> P-marker and a description of the transformation involved.
> Katz and Postal then defined three types of formal
> conditions under which recoverability of deleted elements
> is possible and hence where deletion is permissible :
> (a) If the deleted element is one actually mentioned
> in the structure index of the transformation; (b) if
> the deleted element is strongly identical with another
> element in the same P-marker; and (c) if the deleted
> element is dominated in its underlying P-marker by
> the constituent Pro. Omitting details and simplifying,
> (a) is instanced in the imperative transformation,
> where you is mentioned in the structure index, (b)
> is instanced in reflexive and relative clause transforma-
> tions, and (c) is instanced in sentences like John is
> reading, where in the underlying P-marker the NP
> following the verb is expanded as Pro, this in turn
> is realized as something [or it], and (...) subsequently
> deleted (...) [As Katz and Postal point out,].some
> such underlying form must be posited in order to explain
> the fact that [John is reading] is not infinitely
> ambiguous; it does not mean 'John is reading books
> or magazines or a tombstone, etc.' (Levin 1971,
> 41 and 48).

Conditions (a), (b) and (c) as formulated in the above
quotation are particularly relevant here. As my additions
indicate, (c) can refer either to definite pronominalization
- the situation which is relevant here ('John is reading
IT') - or to indefinite pronominalization ('John is reading
SOMETHING') (cp. Fillmore 1969a, 119).

I shall first consider Levin's first condition. It parallels Jespersen's point about the difficulty of reconstructing some semi-articulate or inarticulate sentences and concerns the possibility of "actually mentioning" certain entities in a structure index. Taking just the example 'A discussion of—', it is easy to realize that there is an indefinite number of elements that could be reconstructed in a structure index for 'A discussion of—', and that there is no way of establishing that one reconstruction is 'better' than another. This is illustrated below in (45) :

<div align="center">(45)</div>

1. Active.
   - (a) There is A DISCUSSION OF— (cp. There is a general examination of— : J 664).
   - (b) This is A DISCUSSION OF—
   - (c) (In this paper/report(etc.) the author gives/presents/includes/offers/etc. A DISCUSSION OF—
   - (d) This paper/report/etc. is/gives/etc. A DISCUSSION OF
   - (e) This short/etc. paper is/gives/etc. A DISCUSSION OF—
   - (f) The major portion of this paper/report/etc. is devoted to A DISCUSSION OF—.
   - (g) Included in this paper is A DISCUSSION OF—

2. Passive.
   - (a) A DISCUSSION OF— is given/presented/etc.
   - (b) In this paper/report/etc. + (a)
   - (c) In this paper/etc. on — + (a)
   - (d) (a) + by the author/the writer/etc.
   - (e) A DISCUSSION OF— is briefly presented/is given in the n pages of this paper/etc.

Surely, there are less possibilities of reconstruction
in the attested sentences quoted below in (46) (a), (b),
and (c), as is suggested in (d) to (g). But the very possibili
ty of having conjunctions of semantically different nominals
such as those underlined in (a), (b) and (c) suggests that
a linguistic theory should generate such nominals directly
in the base <u>as they appear in the surface constructions</u>,
rather than derive them from more complete or explicit
deep-structure representations.

<p align="center">(46)</p>

(a) Introduction to mechanical documentation methods
      Brief <u>description and applications of</u> peek-
a-boo and edge- punched cards... (J 1080).

(b) Experiences with edge notched cards in a neurologic-
al clinic (title)
<u>Description and use of</u> a simplified two-edged
notched card for recording case histories... (J 327).

(c) The authors present <u>a design and discussion of</u>
bibliographic punched card... (J 389).

(d) The author presents/gives <u>a description and the
(a ?) use of</u> ...

(e) <u>A description and the (a ?) use of</u> ... is given/
presented.

(f) ✗ This is <u>a description and use of</u> ...

(g) ✗ The author/the paper includes/offers <u>a discussion
and use of</u> ...

Note, moreover, that the following sentences are
dubious if not completely unacceptable, which to me is a
further argument for the approach I propose :

(47)

(a) ? <u>Discusses</u> x and <u>the paper describes</u> y.
(b) ? <u>Discusses</u> x and <u>the author describes</u> y.
(c) ? <u>A discussion of</u> x and <u>the paper describes</u> y.
(d) ? <u>A discussion of</u> x and <u>the author describes</u> y.
(e) ? x <u>is discussed</u> and <u>the paper describes</u> y.
(f) ? x <u>is discussed</u> and <u>the author describes</u> y.
(g) ? x <u>is discussed</u> and <u>a description of</u> y.

Lastly note that the need for 'incomplete' deep-
structures is also suggested by the use of pronouns in the
following, articulate, sentences :

(48)

(a) The <u>author</u> describes <u>his</u> as yet incomplete work on...
He explains and justifies <u>his</u> method... and
expresses <u>his</u> confiddnce that--- (J 126).

(b) The <u>paper</u> describes incomplete work on ... <u>It</u>
explains and justifies <u>the</u> method... and expresses
∅ confidence that---

(c) ⋇ The paper describes (<u>his</u>) as yet incomplete
work on ...
He explains and justifies <u>his</u> method... and
expresses <u>his</u> confidence that---

In connection with Levin's third point, it also seems
impossible to posit definite pronouns for Agent ('AUTHOR')
and Locative ('TEXT') in the semi-articulate sentences
under discussion. This, at least, is the case with verbal
constructions, as is shown in (49). On the other hand, as
far as pronominalization is concerned, nominal constructions
like '<u>A discussion of</u>---' act as synonyms of nouns like
'<u>text</u>', as is shown in (50).

(49)

(a) ⋇ <u>Discusses</u> ... HE describes...
(b) ⋇ <u>Discusses</u> ... IT describes...
(c) ⋇ ... <u>is discussed</u>. HE/IT describes.

(50)

(a) An evaluation of——. **It** recommends that—— (J 55).
(a') An annual review of——. In addition, **it** includes
    lists of—— (J 230).
(b) Describes—— **This** includes both review of—— and ——
    (J 267).
(c) **A** general description of——. **The description** is
    rather short—— (J 928).
(d) Summary of——, including description of—— (J 31).
(d') An appendix follows, listing—— (J 135).

Let us now turn to inarticulate sentences. Most titles
of the Jannske corpus are of the inarticulate type. But
occasionally, as in the following examples, the first (as
well as other sentences of the abstracts are inarticulate
and consist of just a noun-phrase (cp. Jespersen's "Thanks");

(51)

(a) Searching medicinal chemical literature (title).
    <u>Three representative problems</u> of a manufacturer
    of fine chemicals in searching for information
    in the medical and chemical literautre : ... (J 3).
(b) Mechanized indexing of information on chemical
    compounds in plants (title).
    <u>A system devised for</u> ... (. 168).
(c) Card-form publications in medicine and related
    fields (title).
    <u>Advantages and disadvantages of</u> card-form publica-
    tions. (J 324).
(d) Symbolic shorthand system for medicine and physiolo-
    gy (title).
    <u>A system for</u> classifying and indexing medical
    literature using ... (J 884).

Here again, it seems impossible to decide which
articulate sentence could be said to correspond to the
'best' deep structure representation of the inarticulate
sentences in question (Levin's first point and Jespersen's
observation about "<u>Special edition !</u>", etc.), even though
some reconstructions are definitely ruled out. This is
shown in (52) :

(52)

- The author discusses
- The author presents a
  thorough discussion of
- The author outlines
- The author presents
- The author analyzes
- The author evaluates
  (a method/system)(?)
- The author defines
- The author introduces(?)
- The author stresses
  (advantages)
- The author states
  (problems,advantages)
- The author explores(?)
- etc.
- ✗ The author advocates
- ✗ The author predicts

- three representative problems
  of ...
- a system devised for...
- advantages and disadvantages
  of ...
- a system for ...

The following examples (53), however, suggest that what has
just been said needs to be qualified. In (a), we have
'<u>Study of</u>' introducing '<u>a method for—</u>' in the title, but
only the noun-phrase introduced by '<u>Study of</u>' is repeated
in the first sentence of the abstract. The converse situation,
which is not exemplified, is more frequent : s⌐ (44).
(b) and (c) are even more intriguing. In (b) '—<u>are ALSO
discussed</u>' appears in the fifth sentence of the abstract
WITHOUT any previous occurrence of '<u>discuss</u>' or the like,
and in (c) '<u>these proposals</u>' also occurs without previous
occurrence of '<u>propose</u>' or the like in a preceding sentence.
Example (d) suggests a further problem, namely how inarticul-
ate sentences like '<u>A method for ...</u>' in (a) relate to
ARTICULATE sentences with NO verb such as '<u>discuss</u>' or a
cognate nominal.

(53)

(a) (Analytical and experimental) <u>study of</u> a method
for literature search in abstracting journals
(= title).
A method for literature search using abstracting
journals developed by ... (= first sentence : J 156).

(b) The book-form index requires no searching machine,
and the coordinate searching can be done by a cler-
ical person. Each document is entered in one place,
and the documents can be almost any order. A
document entry consists of a single (or multi-)
character abbreviation for each term the document
possesses, and a document identification. The docu-
ment identification may be anything from a serial
number to a brief abstract; micro-reduced document
identifications <u>are also discussed</u>. (J 721).

(c) The author has developed a set of elements which
should appear in all agreements or contracts between
organizations supporting research and organizations
receiving the support. <u>These proposals</u> are based
on 25 case studies (J 35 .

(d) A system has been designed for preparing biblio-
graphic reports ... (first sentence of J 362).

Considering the above examples I would propose the
following :

(i) Contrary to what I have argued about 'AUTHOR' and 'TEXT',
I feel that it is both possible and necessary to posit
a predicate such as '<u>Study of</u>' or '<u>Discussion of</u>' for
the inarticulate sentences (such as that in 53a), and for
most titles, and articulate sentences that do not contain
such a predicate in their surface form.

(ii) Such sentences could then be treated as instances of
<u>recoverable</u> deletion, provided the <u>appropriate</u> predicate
can be "actually mentioned" in their structure index.
While this no doubt presents problems in examples
like (53) (c) ('<u>These proposals...</u>'), I suggest that
obvious candidates as appropriate predicates are the
most general predicates such as DISCUSS, CONCERN, and
STATE (that). By 'most general', I mean those that can

always serve to replace others on the basis of unilateral
implication rules : thus 'analyze x', or 'explain x', or
'list x' all imply DISCUSS x, but not vice versa.

(iii) If (i) and (ii) are correct, both inarticulate
sentences such as (53) (a) ('A method for ...') and
articulate sentences such as (53) (d) ('A system has been
designed for...') can be represented in deep structure
as having the form Predicate + Topic : respectively
'Discusses (or A discussion of) + Topic = a method for...'
and 'States that + Topic = a system has been designed
for .... On this basis, such sentences can be related by
implication rules to articulate sentences of the type
'The author discusses x'(cp. implicit performatives).

(iv) While I realize that the notion of topic in linguistic
theory is far from clear, I suggest that my proposal
to give the same semantic analysis (Predicate + Topic)
to syntactically different verbs like 'discuss' and
'state that' is not unreasonable considering the possi-
bility of paraphrases illustrated below in (54); (d)
is based on a suggestion made by Lakoff (George Lakoff
1969c, 30-31).

(54)

(a) Original sentence (sentence 1 of J 404).
Before any automatic system for storing, searching,
and retrieval of information can be effective, it
is necessary to develop an efficient and economic
method of placing and locating the information in
the system.

(b) Modified version 1.
STATES THAT + (a) = Topic

(c) Modified version 2.
DISCUSSES + Topic = the need to develop an efficient
and economic method of placing and locating (the)
information (in the system) before any automatic
system for storing, searching, and retrieval of
information can be effective.

(d) Modified version 3.
CONCERNING + Topic = automatic systems for ...
IT IS STATED that + Topic = it is necessary to
develop an efficient and economic method of placing
and locating the information in THE...

355.

(v) If (iv) is correct, it suggests that sentences like
(54) (a) <u>as a whole</u> can be considered as a Topic. This,
I suggest, is not unreasonable and can be regarded as
a formulation of the 'implicit performative hypothesis'
in terms of a case theory framework. Under such a view,
then, there would be no difference between (54) (a)
and (b) as far as their Topic is concerned, the topic
of (54) (c) would be a noun-phrase with '<u>need</u>' as a
head-noun, and (54) (d) would have two Topics :
'<u>automatic systems for...</u>', and a that-clause which is
equivalent to (54) (a) and (b). Similarly, the only
difference as far as Topic is concerned, between
"(STATES that) <u>a system has been designed for ...</u>"(J 362)
and, say, "(DISCUSSES) <u>a system devised for...</u>" (J 168)
would be captured in the grammar, in terms of the
difference between a sentence and the corresponding noun-
phrase with a relative clause. In the next chapter, I
shall argue that the above proposals are relevant to some
problems of identity in discourse and of discourse
structure.

3.3. CONJUNCTION, RELATIVE CLAUSES AND PRESUPPOSITION.

3.3.1. Some marks and factors of identity.

Ignoring various details and problems, the review of recent studies on conjunction presented in 3.1.4. leaves us with the general assumption that and-conjunction imposes identity constraints on what it conjoins, and with the following, more specific, assumptions :

(i) sentences in a text or discourse are relatable by and or some other conjunction, so that some identities must be statable between them (Katz and Fodor, and Harris);

(ii) identities must be statable between conjoined sentence constituents, such as noun-phrases or verb-phrases (particularly under the hypothesis that the source is a conjunction of sentences);

(iii) it is necessary to distinguish between symmetric and asymmetric 'and' and these two types of 'and' may require different semantic representations.

In this section, I shall briefly discuss and illustrate three factors of identity :

(i) case structure;

(ii) marks of time sequence;

(iii) deixis.

Later in this chapter, I shall examine some of these points (next to a few others) in more detail. To begin with, I shall suggest briefly how case representations can throw light on the problem of identity in conjoined sentences. In (1), (2), and (3), I wish to suggest the following. On the basis of the hypotheses put forward in the preceding chapter, the sentences of the illustrative abstracts quoted in (a) can be represented in terms of the general format <u>Predicate + Topic</u> given in (b).The representations of (b) 'normalize' the texts, on the basis of an application relation between various sentences and DISCUSS (or STATE) + Topic.

in such a way that each sentence has the same representation; and, on the basis of both the Harris-Lakoff hypothesis about identity between conjoined sentences and the Katz-Postal hypothesis about texts being reducible to a conjunction of their constituent sentences, we can represent all the illustrative abstracts as <u>Predicate + Topic AND Predicate + Topic AND ...</u>

(1)

(a) <u>Points out</u> the value of abstracts (particularly informative abstracts), <u>sets forth</u> standards for a good abstract and <u>examines critically</u> a number of abstract services and <u>shows</u> why they do not meet the standards!
<u>Discusses</u> the question of who should prepare abstracts and indexes and <u>examines critically</u> Australian abstracting and indexing service (J 399).

(b) <u>DISCUSS + Topic</u> ("the value..."), <u>DISCUSS + Topic</u> ("standards...") and <u>DISCUSS + Topic</u> ("a number...") and <u>DISCUSS + Topic</u> (the reason "why ...").
<u>DISCUSS + Topic</u> ("the question...") and <u>DISCUSS + Topic</u> ("Australian...").

(2)

(a) Methods used by the Thermophysical Properties Research Center for handling collected bibliographical information on thermophysical properties <u>are described</u>. <u>Details are given regarding</u> the method of coding abstracts for reproduction on IBM cards, particularly the coding of substances and the system of substance classification and notation used. Procedures used to store the punched card information on magnetic tape <u>are described</u>. Systems of retrieval of this stored information <u>are delineated including a description of</u> the organization of the retrieved bibliography and of the machine techniquesinvolved. The system <u>described</u> is claimed to be extensible to some other portions of scientific literature. <u>Announced in this paper are</u> the Center's plans to make available an annual index to the literature of thermophysical properties, expected to start publication in 1959. This book-form index--essentially the output of the Center's retrieval equipment--<u>is described</u> in prospect as a definitive, cross-indexed bibliography of every paper, book, report, or patent containing thermal property information. (J 985).

(b) <u>DISCUSS + Topic</u> ("methods used..."). <u>DISCUSS + Topic</u> ("the method of coding..."). <u>DISCUSS + Topic</u> ("procedures used..."). <u>DISCUSS + Topic</u> ("systems of retrieval...") and (?) <u>DISCUSS + Topic</u> ("the organiza- tion of ..."). <u>SAY ("is claimed") + Topic</u> ("the system described is claimed..." : I shall return below to the problem arising with "system described"). <u>DISCUSS + Topic</u> ("this book-form index... as ..."; I do not go into the problem raised by "describe x <u>as</u> y").

(3)

(a) Information retrieval techniques to make the vast store of knowledge conveniently accessible are still far short of what's needed. The difficult problem is not devices and technology, it's the intellectual job of matching what 's known and what's wanted (J 4C7).

(c) <u>STATE + Topic</u> (that "information retrieval techniques... ..."). <u>STATE + Topic</u> (that "the difficult problem...").

Another way in which case representations can serve to state identities is the following. As will be remembered, Fillmore's first definition of the "<u>Objective</u>" case was : "the semantically most neutral case, the case of anything representable by a noun whose role in the action or state identified by the verb is identified by the semantic interpreta- tion of the verb itself" (F 1. e 1968a, 25). Later, he defined the "<u>Object</u>" as "the entity that moves or changes or whose position or existence is in consideration" (1969a, 77). Both definitions can serve as a basis for regarding the NP's with '<u>information</u>' in (4) (a) as exponents of the Objective case so that part of the sentence given in (b) can be represented by a conjunction of <u>N + 0 = information</u> as in (c) (following Stockwell <u>et al.</u> 1968, PS rule 7, 34). Without going into the problem of 'nominalization', it is clear that the conjunction of gerunds in (4) is equivalent to a conjunction of sentences.

(4)

(a) Most information retrieval systems depend upon extracting information from documents, processing such information in some manner and finally placing this information in a file so that the information can be retrieved by machine (J 321).

(b) Extracting information, processing such information, and placing this information.

(c) <u>N + O = information</u> AND <u>N + O = information</u> AND <u>N + O = information</u>.

I suggest that such an analysis can serve to explain why the Lévy concordance is entitled "Etude de concordance interlexicale <u>dans le domaine de l'information scientifique et technique</u>" (<u>in the field of scientific and technical information</u>) Under my assumption, it seems natural that the authors should have thought of <u>scientific and technical information</u> as the 'topic' (in R. Lakoff's sense) common to all their classification entries. This may also serve to explain why <u>scientific information</u> often serves as a synonym of '<u>documentation</u>'.

A less obvious way in which case representations can throw light on problems of identity in conjunction has to do with the <u>Result</u> or <u>Goal</u> case. Fillmore defines <u>Result</u> as "the entity that comes into existence as a result of the action". Now, compare the conjunction given in (5) (a) and the near-synonymous expression in (b), and note the oddity of (c) :

(5)

(a) <u>develop</u> and <u>apply</u> automatic techniques (J 819).
(b) the system was <u>developed</u> for <u>use</u>... (J 115).
(c) ? apply (use) and develop a system.

(b) suggests that (a) is an instance of <u>asymmetric</u> conjunction.
(more on this in 3.3.3.). Now, <u>develop</u> can be analyzed as
requiring a <u>Result</u>, whose exponents are '<u>techniques</u>' in (a)
and <u>system</u> in (b), while, the same words in relation to '<u>apply</u>'
and '<u>use</u>' must be regarded as exponents of another case -
presumably <u>Instrumental</u>.It is therefore reasonable to assume
that part of the meaning of asymmetric conjunctions could
be represented in terms of a shift in which the <u>same entity</u>
is a Result or Goal in the first conjoined sentence and
another case in the second. The same turns out to be true
of an example   given by Langendoen (1969, 45) :

<div align="center">(6)</div>

(a) The bartender made the martini,
    and then drank it.
(a') MAKE + Goal = martini <u>and then</u> DRINK + Object =
    martini.
(b) x The bartender drank the martini, and then he made it.

One last point may be made here about the relations between
case structure and conjunction. It is clear that for somebody
to be able to draw some '<u>conclusion</u>', there must be some
'<u>basis</u>' for the conclusion. Now, the Janaske corpus provides
ample evidence for considering that some constituent, which
I would represent as an Instrument, can be posited to be
present in all sentences with '<u>conclude</u>' and a few other
verbs : 'Conclude + Instrument = on this basis'. The instru-
mental analysis of these constituents is justifiable on the
basis of Fillmore's definition of the Instrumental as "the case
of the inanimate force or object causally involved in the
action or state identified by the verb" (1968b, 24) or of the
Instrument as "the stimulus or immediate physical cause of an
event" (1969a 77). In (7) (a), I give an example in which no
Instrumental constituent is present. Such a constituent, however
can be assumed to be implicit, on the basis of the other
examples given in (7). The constituents which I assume to be

Instrumentals are underlined, and the relevant predicates are
capitalized; the question whether the underlined expression
in (e) is a constituent of the sentence with 'reviews' is of
course debatable, and current research (to my knowledge)
provides no undisputed answer to the question (Fillmore 1969a
does not discuss examples like those given below; nor does
he posit Instrument in the lexical entry of 'verbs of
judging' such as CRITICIZE).

<div align="center">(7)</div>

(a) In addition to presenting a general review of——,
and a review of——, this article presents a review
of——. The author CONCLUDES that—— (J 5).

(b) Testimony before the Moss Committee——. It is CLEAR
from the testimony that——. All the testimony POINTS
to the CONCLUSION that —— (J 784).

(c) The author's belief is that——. Upon this foundation,
he CRITICIZES—— (J 54).

(d) On the premise that——, the writer DISCUSSES—— (J 289).

(e) Based on——, this paper REVIEWS—— (J 336).

I suggest, then, that such sentences can be represented as
'Predicate + Instrument', where Instrument is a sentence
or refers to the preceding sentences in the text, as does
"Upon this foundation". If I am correct,   Instrumental
NP's like "Upon this foundation" serve to introduce identity
into a text, and we can represent texts like (7) (c) as
follows : S1 and S2 (= CRITICIZES + I ⬝ NP = "On the basis
of S1").

Marks of time sequence are the second factor of identity
I wish to examine here. I shall not further discuss the
problem in this thesis, but I think that it is worth mention-
ing, even though no linguistic theory known to me has any-
thing to say about the expressions exemplified below.
First, there are various expressions which serve to refer
to aspects of the linearity of discourse (generally the
abstracted document). Superficially they can belong to

various parts of speech categories but they have comparable
meanings. Consider the following attested expressions and
their acceptable and unacceptable paraphrases in (8) :

(8)

(a) A brief review of—. —is tabulated. The article
  <u>ends</u> with a plea for—. (J 487).
(a') ... <u>Finally</u>, there is a plea for—
(a'') ... There is a <u>final</u> plea for—
(a''') ...x There is a <u>last</u> plea for—
(b) On the premise that—, the writer discusses—
  <u>In the last</u>, he cautions against— (J 284).
(b') ... The article <u>ends</u> with the warning that— (?)
(b'') ... There is a <u>final</u> warning that— (?)
(b''') ...x There is a <u>last</u> warning that— (?)
(c) This report deals <u>with</u>—.— are given;...
  A <u>final</u> chapter is devoted to— (J 872).
(c') ... <u>Finally</u>, there is a chapter— (this can, but
  need not be equivalent to c)
(c'') ...A/The <u>last</u> chapter is devoted to—
(c''') ... The report <u>ends</u> with a chapter...
(c'''') ... <u>In the last</u>, there is a chapter...
(d) In addition to describing—, the author describes—.
  He also includes—.
  <u>Finally</u>, there is presented an explanation of—
  (J 421).
(d') x There is a <u>final</u> explanation of— (?)
(d'') x There is a <u>last</u> explanation of—
(d''') <u>In the last</u>, there is an explanation of—
(d'''') The report/article/etc. <u>ends</u> with an explanation
  of...
(e) ...some questions are raised, among them : (...)
  <u>And finally</u>, will ... be ... ? (J 374).
(e') A/One <u>final/last</u> question is whether ... will...
  be...
(e'') The author/The paper (?) <u>ends</u> by asking the
  question whether...
(e''') <u>In the last</u>, the question is asked whether...

On the basis of (e), we can assume that an '<u>and</u>' is implicit
in all the other sentences given in (8). An argument for
regarding this '<u>and</u>' as asymmetric is that the above sentences
would be very odd indeed at the beginning of a text.

The other expressions which refer to the linearity of discourse and which seem to be naturally linked by (or equivalent with) asymmetric _and_ are enumerations and expressions with _before_, _after_ or equivalent verbs as illustrated below in (9). Note that the last point of the enumerations could be introduced by _and finally_, or various synonyms. Also note the equivalence between '_Also_' standing at the beginning of a sentence in (e), and the sentences with '_follow_' and '_is followed_'. The only expressions which are not clearly asymmetric are '_in addition_'/'_in addition to_'/'_Additional_'... in (g), (k), and (1)

(9)

(a) ·—is divided into the following parts :
(1)—; (2)—; (3)—; and (4)—.
(b) _First_, this is a discussion of—. _Second_, the consequences of— are drawn. _Third_, this describes— (J 853).
(c) A description of— _is followed_ by two appendices... (J 33).
(d) This annotated bibliography of— presents—. An appendix _follows_, listing—, with a brief description of—. — indexes _follow_ appendix.(J 135).
(e) A review of—. The author concludes that—. _Also_,— . As a final comment the author _adds_ that—
(f) _After_...
(g) The authors describe—. _Preceding_ discussion of—, a brief description is given of—. A short discussion of— _terminates_ the paper. (J 638).
(h) ... A bibliography of— is _appended_. (J 265).
(i) —— is traced—. A brief evaluation is _then_ given of— (J 880).
(j) _In addition_, it appears that ...(J22)
(k) _In addition_ to presenting a general review of—, and a review of—, this article presents a review of—. (J 5).
(1) _Additional_ information concerns— (J 7).

I have no idea how linguistic theory could reconcile the superficial diversity of these expressions with their semantic similarities. The problem is even worse if one considers that the function of many if not most of the expressions given above is by no means restricted to that of marking discourse linearity. This is why the term,

'_time sequence_' has been used above in connection with such expressions. Consider the attested occurrence of _and finally_ in (10) (a), together with some possible paraphrases.

(10)

(a) Most mechanized information retrieval systems depend upon extracting informati n from documents, processing such information in some manner _and finally_ placing this information in a file so that the information can be retrieved by machine. (J 321).

(b) ... depend upon the following sequence of operations ... _After_ being processed, the information is placed in a file...

(c) The processing of the information _is followed by_ its placement in a file...

(d) The information is extracted ... (and) _then_ ...

(e) The _last_ operation is the placement... _so that..._

The paraphrases of '_and finally_' in (10) suggest that this '_and_' is of the asymmetric type, and that Robin Lakoff's hypothesis is correct when she advances that what one means by asymmetric conjunction, and common topic in connection with it, is 'possibility of reduction' to an asymmetric predicate (such as '_follow_' : R. Lakoff 1970, 27). On the other hand, Robin Lakoff's conjecture captures only part of the problems of asymmetric '_and_', as I now wish to show by examining one last mark of identity, namely _deixis_. I shall do this by again returning to the sentence quoted in (4) (a) and (10) (a). The points I wish to make in connection with deixis are illustrated in (11). In (11) (a) the sentence given in (10) (a) is repeated with the deictic expressions underlined. It seems to me that (11) (b) which preserves these deictic expressions but which has the form of a discourse is an acceptable paraphrase of (a). (c) shows that the discourse of (b) can be converted into a single sentence with relative clauses. (d) shows that (b) can be paraphrased by a discourse in which the deictic expressions are replaced by NP's with relative clauses or relatable constructions. The problem is of course why and when

such conversions are possible, and how to represent the process. I shall return to this question below.

(11)

(a) Most information retrieval systems depend upon extracting information from documents, processing such information in some manner and finally placing this information in a file so that the information can be retrieved by machine (J 321).

(b) Most information retrieval systems depend upon three main operations.(The) information is extracted from (the) documents. Then this/the information/it is processed in some manner. (And) finally this/ the information/it is placed in a file. As a result/ In this fashion, the information can be retrieved by machine.

(c) Most information retrieval systems depend upon extracting from the documents some information which is then processed in some manner and which is finally placed in a file and which can therefore/ then be retrieved by machine.

(d) Most information retrieval systems depend upon three operations. (The) information is extracted from documents. Then the information (which has been) extracted (from documents)/the extracted information is processed in some manner. Finally the information (which has been extracted... AND) which has been processed/the processed information is placed in a file. The information (which has been extracted,... processed... AND) placed in a file/the stored information is then retrieved by machine.

The following examples given in (12) suggest further complexities in the problem of deixis : (a) and (b) suggest that some reference to time sequence may be necessary ("previously"), in addition to some reference to tense ("which has been previously classified") and to other abstract notions such as 'result' ("resulting"); (c) illustrates the problem, which is a well-attested one in the corpus under investigation, of deictic NP's with no surface antecedent; (d) illustrates the converse situation in which an NP is relatable to a following sentence :

(12)

(a) Encoding of information <u>previously</u> classified.
(J 65/cp. J 511).

(a') Cp. : encoding of <u>the(this)</u> information (11a).

(b) ... certain documentation services... involve
a regular scrutiny of over one thousand current
periodicals <u>with consequent</u> indexing and distribu-
tion of the <u>resulting</u> information. (J 466).

(b') Cp. : and indexing and distribution of <u>the(this)</u>
information.

(c) Describes a new experimental attempt to develop
automatic classifications of chemical concepts
and interrelations in the disclosures of the
patent literature. <u>In this system</u>, the classifica-
tion system is made... (J 19).

(d) The system described employs three established
<u>techniques</u> : <u>microfilming</u>, punched-cards, and
xerography. Documents to be held on file are
<u>microfilmed</u>. (J 35).

3.3.2. Discourse, relative clauses and lexical items.

As appeared most clearly in the quotation made from
Kuroda in 3.1.4., the question of discourse has long been
a neglected area in TG linguistics, and, with the notable
exception of Harris, this is also true of other trends of
linguistics. Things have changed, however, at least if
Robin Lakoff's approach to sentence conjunction may be
regarded as typical of an evolution in TG studies. There is
further evidence for such an evolution in the fact that
in recent years several TG linguists independently arrived
at the conclusion that the theory should account for the
paraphrase relations between discourses and sentences contain-
ing a noun modified by a relative clause (restrictive or not).
These relations have been most systematically examined by
Annear Thompson in an unpublished dissertation and in two
papers (1970 a and b), and, on the basis of her work, Langen-
doen has proposed rules which optionally derive a sentence
containing a noun modified by a relative clause from "a
discourse consisting of two sentences" (Langendoen 1970,
141-150). Thus, the application of Langendoen's rules has
the effect of replacing optionally a pair of sentences such
as 'My cousin married a girl. She became pregnant' by 'The
girl whom my cousin married became pregnant'. Note that
in Langendoen's account: (a) only pairs of coordinate
sentences are considered as a discourse source; (b)
the problem of identity which must hold between them is
reduced to the (equally unsolved) problem of 'pronominaliza-
tion' ('a girl. She...'); (c) the nature of the relation
between the two coordinate sentences is not specified (it
is just represented as a period); (d) the (optional) applica-
tion of the rules for relativization has the effect of
deleting the discourse source. Points (a) and (c) are also
true of Annear's account; (d) is irrelevant, since Annear

does not propose any transformations. In particular, Annear
leaves open the question of determiners (b). She argues
that this has to do with presumably non-syntactic questions
concerning "certain presuppositions which the speaker makes
about the extent of his listener's knowledge", and she
proposes discourse sources for relative clauses of the type
(I met girl) (girl speaks Basque). For Annear, then, the
identity (b) which must hold between the pair of sentences
serving as the source of a sentence with relative clause
seems to be a matter of lexical identity (Annear 1970b, 43).

In spite of their lack of elaboration, the Annear-
Langendoen proposals represent a revolution in TG linguistics,
since there now exist basically two types of hypotheses
about relative clause formation, and there is (as far as I
can see) no sign that a reconciliation between them will
eventually emerge. Next to the Annear-Langendoen hypothesis,
which regards discourse as basic and relative clauses as
derived by optional transformations, there are the various
hypotheses initiated by Chomsky according to which relative
clause formation is obligatory, given a deep structure
configuration consisting of a noun-phrase with one or more
embedded sentences or clauses. As Annear points out (1970b,
42), there is "virtually no agreement" among those who propose
this type of analysis as to its detailed formulation (see
in particular the review of these hypotheses in Annear op.cit.
and in Stockwell et. al. 1968, Ch. VIII).

In this section, I shall argue that it may not be
necessary to reconcile the Annear-Langendoen proposals with
the others, because there are reasons for believing that
BOTH discourse and relative clauses are basic. I shall
do this around the five points (a-e) which I have just
mentioned concerning the lack of elaboration of the Annear-
Langendoen hypothesis.

I shall go out from texts, that is, abstracts rather than from discourses consisting of no more than two sentences (a). I shall examine the nature of the identity which can be posited to hold between sentences of an abstract. As I wish to show, this identity does not _always_ manifest itself by superficial marks such as deixis ('a _girl_... _This_ _girl_') and/or matching lexical items (cp. Harris's "word repetition" : '_girl_..._girl_') (b). I shall also discuss the problem of what can be said to underlie the _period_ of discourse (e.g. '_and_', or some other conjunction : c), and I shall present arguments against the deletion of the discourse source - one of the effects of the rules of relativization in Langendoen's formulation (d). I shall present successively evidence FOR and AGAINST the Annear-Langendoen hypothesis.

A most typical feature of the abstracts under investigation is the occurrence in the second or following sentences of deictic noun-phrases (for instance, _the_ N or _this_ N, next to various other possible constructions) where N already occurred in a preceding sentence. When this is the case, it is often possible to convert the abstract into a single sentence with relative clauses, changing the deictic noun-phrase into a relative pronoun. The deictic noun-phrase, then, appears as one of the possible superficial marks in discourse of the identity required for (optional) relative clause formation. As the following examples suggest (13-15), the conversion into a single sentence with relative clauses requires only a few grammatical and occasionally lexical changes (I shall not discuss these here, as I wish to discuss similar problems later in this section). The converse change from a single sentence to a discourse is exemplified in (16). Also note that (13) suggests the need for relative clause formation to be _optional_ : while a single sentence containing, say, ten

thousand relative clauses is conceivable, people usually
prefer to express themselves by means of the corresponding
discourse - a fact which still awaits explanation, within
a theory of performance, in psycholinguistics, or otherwise.

(13)

(a) A system has been designed for preparing bibliograph-
ic reports using an automatic typewriter, the
"Flexowriter", which prepares a punched paper
tape simultaneously with the original typing of
the reference cards during the progress of the
literature search. These tapes, made on a prefolded
8 1/2'' fold tape, may be cut apart, filed, and
rearranged in any desired order for the final
report... The use of this system has resulted in
a saving of 60 % over the time required for manual
typing of a bibliographic report. (J 362).

(b) A system (,the use of) which has resulted in a
saving of 60 % (...) has been designed for preparing
bibliographic reports using an automatic typewriter,
the "Flexowriter", which, simultaneously with (...),
prepares a punched paper tape, which is made on a
prefolded 8 1/2'' fold tape and may be ...

(14)

(a) The ever increasing need for a more efficient
organization of medical information in the Soviet
Union indicates that the present decentralized and
inefficient system of medical documentation should
be replaced with a centralized system. The central-
ization of medical documentation would permit a
better utilization of means and personnel already
available to various organizations engaged in med-
ical documentation and better dissemination of in-
formation through improvement in methods of pres-
entation and enlargement of the scope of medical
documentation. (J 42).

(b) ... with a centralized system which would permit...

(15)

(a) The author attempts to demonstrate the existence of
a theory more basic than that now published in the
literature of the field, and to show how a more
fundamental theory might lead to conclusions hither-
to not realized. This basic theory is based on the
author's definitions of "nature" and "information"
which were written to meet the needs of the pro-
fessional bibliographic organizer. (J 673).

(b) ... how a more fundamental theory which is based
on... might lead to ...

(16)

(a) proposes the creation of national <u>documentation</u>
<u>centers which</u> would cooperate with one another
in the dissemination and indexing of sciontific
literature (J 200).
(b) Proposes the creation of <u>national documentation</u>
<u>centers. The/these/such centers/They</u> would...

When the sentences of an abstract have a verb of the
type discussed in 3.2. as main verb, as in (17) (a),
the sentences can be conjoined by '<u>and</u>' as in (b) and this
conjunction can in turn be converted into a single sentence
with a relative clause, as in (c) :

(17)

(a) <u>Describes</u> the Uniterm **system of indexing** being used
by the Civil Engineering Department of Case Insti-
tute of Technology. The advantages and the limi-
tations to <u>this indexing system are given</u>. (J 674).
(b) Describes the Uniterm system... <u>and gives</u> the
advantages and limitations to (?) this indexing
system.
(c) Describes <u>the Uniterm system</u>..., the advantages...
<u>of which</u> are given.

My discussion in 3.2.3. offers a basis for predicting some
changes required by such cohversions : o.g. "are given"/
"gives".

Langendoon himself provides an argument (which is
based on Ross 1967) for his treatment of discourse as
basic. One of the situations defined by Ross in which
relativization is <u>ruled out</u> is that in which the relative
pronoun would have to originate in a relative clause which
is itself inside a relative clause. This is exemplified
below in (18); compare the unacceptable paraphrase in (b)
with the acceptable one in (c) :

(18)

(a) A general description of the <u>Japan Information
Center of Science and Technology which</u> was es-
tablished for the purpose of contributing to the
development of science and technology by collect-
ing, processing, storing, and retrieving scientific
information. At the time of this report, <u>the Center</u>
was planning to change over to a mechanized storage
and retrieval system similar to that used by the U.S.
Patent Office. Services available from <u>the Center</u>
and the charges for each are listed. (J 1).

(b) A general description of the J IC Sc T which was
established for the purpose of contributing to the
development of science and technology by ..., (and)
<u>which</u> at the time of this report was planning to
change over to..., and <u>x which</u> the services THAT
ARE AVAILABLE FROM are listed.

(c) A general description of the J IC Sc T <u>which</u> was
established..., <u>which</u> at the time... and <u>the services
of which</u> are listed.

A further argument for the Annear-Langendoen hypothesis
is the following. Some abstracts cannot be converted into a
single sentence with relative clauses WITHOUT A CHANGE IN
MEANING. The reason seems to be that the period of written
text cannot be paraphrased by '<u>and</u>' without changing the
meaning. The meaning - preserving paraphrases are with
some <u>other</u> coordinating conjunction, or some equivalent
subordinating construction. Thus, in (19), I would say that
paraphrases (b) and (b') do not preserve the meaning of (a)
as do (c) and (d).

(19)

(a) A Uniterm system is used by Humble Oil... It has
proven to be most satisfactory for their purposes.
<u>The system</u> cut indexing time, increased retrieval
speed,...

(b) The Uniterm system used by H.O. has proven to l most
satisfactory for their purposes <u>and it</u> cut indexing
time,...

(b') The Uniterm system used by H.O. <u>which</u> cut indexing
time, increased retrieval speed, (...) has proven
to be most satisfactory for their purposes.

(c) The Uniterm system used by H.O. has proven to be
most satisfactory for their purposes <u>in that it</u> ...

(d) The Uniterm system used by H.O. has proven to be
most satisfactory for their purposes. <u>The reason is
that</u>...

Occasionally, in some abstracts, there is no surface
mark of identity - such as deictic noun-phrases - of the
type required for relative clause formation. Another, similar,
situation is that in which deictic noun-phrases occur without
any previous occurrence of a lexically identical antecedent.
This could at first be regarded as an argument for the Annear-
Langendoen conception of discourse as basic. On the other
hand, it seems that it is always possible to reconstruct the
texts in such a way that relative clause formation is possible.
Unfortunately, no theory is currently able to explain such
reconstructions; I shall return to some of them later in this
thesis. The texts, together with the corresponding reconstruc-
tions and relative clause constructions, are given below
in (20) to (25) :

(20)

(a) Proposal for the creation of a permanent international
organization which would have charge of the accumula-
tion of medical data in the world literature, and of
an analysis and synthesis of these data. By means
of cards and leaflets, physicians could have informa-
tion on new developments in all subjects and fields
of medical science (J 432).

(b) Proposal for the creation of a permanent international
organization which would have charge of the accumula-
tion of medical data in the world literature, of an
analysis and synthesis of these data, <u>and of the
distribution of cards and leaflets</u>. By means of <u>these</u>,
...

(c) (as in b except:) <u>by means of which</u>...

(21)

(a) ... this is becoming more and more <u>difficult</u>. The
keys to <u>this problem</u> are... (J 889).

(b) ... this is becoming more and more of a <u>problem</u>/
...this raises more and more <u>problems</u>. <u>The keys to</u>
<u>these problems/this problem</u>...

(c) ..., <u>the keys to which</u> ...

(22)

(a) /. report of a three weeks' <u>trip to 22 technical infor-</u>
<u>mation centers</u> on the continent of Europe. The
three aims of the trip were : 1) to examine methods
of operating information services; 2) to study the
attitude of industrial management and research to-
ward technical information, and 3) to contact some
of the leading people in the technical information
field in Western Europe. <u>Each center visited</u> is fully
described, and its available information and library
services <u>is</u> given. (J 887)(sic).

(b) ... trip for the purpose of <u>visiting</u>... <u>centers</u>...
Each center visited...

(c) /. report of a three weeks' trip (the three aims <u>of</u>
<u>which</u> were : ...) for the purpose of visiting 22
technical information centers on the continent of
Europe, <u>which</u> are fully described and <u>whose</u> available
information and library services are given.

(23)

(a) This paper analyzes—. It also analyzes—. <u>On these</u>
<u>bases</u>, the paper presents— (J 963).

(b) This. paper analyzes— and —, <u>upon which bases</u> it
presents—

(c) This paper analyzes— and —, <u>on the basis of which</u>
analyses it presents—

(24)

(a) The author's belief is "that a clear understanding of
the distinction between information retrieval systems,
the function of which is to make detailed, factual
information available, quickly, cheaply, and re-
liably, and literature search systems, the function of
which is to provide a bibliography of documents for
the use of the investigator of a specific problem, and
of the related distinction between the information-
condensing function of tne abstract and the clue-
providing function an index, is necessary for
progress in information searching". <u>Upon this founda-</u>
<u>tion</u>, he criticises six theories upon which ex-
perimental information retrieval systems have been
based. These six theories are found in the writing
of such well known groups and individuals as Perry
and /.ssociates and Dr. Mortimer Taube. (J 54).

(b) — <u>upon which foundation</u> he...

(c) The author's belief, which serves him as a <u>foundation</u>
for criticizing six theories upon which... and which
are found in ..., is that ...

(25)

(a) Describes a new experimental attempt to develop
automatic classifications of chemical concepts and
interrelations in the disclosures of the patent
literature. In this system, the classification system
is made from the disclosures in the patent literature,
thereby making it possible to... (J 19).
(b) Describes ... in which the classification system is
made from the disclosures...

On the basis of the arguments given so far, there are
good reasons for believing that discourse is basic and that
relative clauses are optionally derived from discourses. I shall
now turn to counter-arguments, which go to show that construc-
tions with or relatable to relative clauses are also basic.

Let us first consider definitions where the definiendum is
a construction usually regarded as relatable to relative clauses,
such as "inverted file" in the following example (26). In such
definitions, it seems that both the discourse paraphrase (b)
and the corresponding relative clause construction are unaccept-
able (Postal has given most striking examples which show that
noun-phrases with relative clauses and nouns synonymous with
them have different syntactic behaviours. Compare, for instance,
'People who smoke really shouldn't do so' and 'x Smokers really
shouldn't do so', 'People who study Iroquoian sometimes deny
they do so' and 'x Iroquoianists sometimes deny they do so' :
Postal 1969, 217) :

(26)

(a) An inverted file is an organizing system for the
stored information in a machine memory so the informa-
tion is divided into as many sections as there are
different codes representing all of the subunits of
information in all of the documents in the memory
of the machine.
(b) x Some files are inverted. These files constitute a
system. This system organizes information. The informa-
tion is stored in a memory. The memory is in a machine.
The information is organized so that...
(c) x A file which is inverted is a system which serves
to organize...

374.

Second, relative clauses with 'so'and 'thus' meaning 'in this
fashion' cannot be derived from a discourse source already
containing 'so' or 'thus'. These adverbials would have to be
introduced transformationally, and, as far as I can see,
there is no justification for doing this. What is true of
NP's with 'Ved' and 'so' or 'thus' is even more obviously
true of NP's with words like "(the) above-mentioned", and
synonyms, as is shown in (27): for instance, you cannot derive
the sentence 'I am interested in the above-mentioned NP" from
"x I mention NP above and I am interested in NP". A similar
example with 'so' is given in (28). In (27), note the occurrence
of "each publication (which is) thus uncovered" and of "the
above-mentioned properties" with NO previous occurrence of
'uncover'or 'mention'; I shall discuss what I believe to be a
similar situation below in connection with example (30).

(27)

A method for literature search using abstracting journals
developed by Lykoudis, Liley and Touloukian. Consists
of searching an abstracting journal for a certain period
of time and then obtaining new information by going
through the bibliography section of the papers located
by the direct search of the abstracting journal. This
procedure can be repeated in cycles. According to the model
of this method a direct search of Chemical Abstracts is
made to locate publications on seven thermophysical
properties (thermal conductivity, specific heat, viscosity,
diffusion coefficient emissivity, thermal diffusivity, and
Prandtl number) for all matter for the period between
January 1954 and December 1967. The references cited by
each publication thus uncovered are searched to obtain
further new information on the above-mentioned properties.
(J 156).

(28)

The essential purpose of literature searching is to
locate those documents within a collection which
have a bearing on a question asked of it. In con-
sidering the design of a system for selecting and
correlating information in response to questions, it
is important to recall that, in the particular appli-
cation of selecting systems, the retrieval of needed
information from a file is usually in two steps :
(1) the selection system is operated to identify those
documents of probable pertinent interest; and (2) the
documents so selected are subjected to personal review
to locate the desired information. (J 773).

Third, contrary to the rules formulated by Langendoen
on the basis of the Annear hypothesis, I have found many
examples in the abstracts of the Janaske corpus in which
the occurrence of a relative clause does not go together
with a deletion of its discourse source. Let me give just one
of these examples. Compare the underlined portions of the
following quotation (29) :

(29)

As an improvement on the card reference information
file commonly used by engineers, a modified **system**
using marginal punched cards **is described**. This
punched card system enables information to be
retrieved from a file on the basis of several classi-
fications. Specifically, **the system described** uses....
(J 440).

This suggests that either the deletion of the source conjunct
of relative clauses should be made **optional** in the Annear-
Langendoen rule, OR that the deletion of the **relative clause**
should be made optional, with both this relative clause and
its discourse sources being generated in the base.

Moreover, it is often the case that the superficial
'discourse source' and the relative clause construction do not
match lexically. The situation with which I am concerned
here (30) is, I believe, different from that of (22) ("**trip**"...
"**visited**"). Here, a reconstruction by adding some material is
neither necessary nor possible. If the sentences with '**deals
with**' and '**is concerned with**' in (30) are rightly regarded
as the source of '(systems **described**)' - as I think they
should be - the abstract quoted below suggests that **lexical
insertion** should be made sensitive to the kind of implication
rules I have proposed in 3.2. : '**TEXT' deals with TOPIC** as
well as '**TEXT is concerned with TOPIC**'imply **TOPIC is
described**, and by virtue of this implication "**Systems described
are...**" (= ... are the systems which are described) can be

376.

used just as if a sentence such as, say, 'punched-card systems <u>are described</u>' actually occurred in what precedes :

(30)

This article is divided into five parts. "The first <u>deals with</u> edge-notched punched cards and related equipment; the second with tabulating type punched cards; the third is <u>concerned with</u> supplements to punched cards; the fourth with other types of equipment related to information processing; and the fifth with ancillary equipment for punched-card <u>systems</u>." <u>Systems described</u> are Keysort, E-Z Sort, Unisort, Findex, Flexisort, Needlesort, Zatocoding, Microtape, Uniterm, and Filmorex. (J 822).

A similar absence of lexical match is exemplified below in (31). Compare "<u>which (= and it) is permanently filed in a central repository</u>" and "<u>on deposit</u>"(= 'which is filed') in the following abstract :

(31)

A letter to the editor proposing the following system of communication research results : The contributor prepares a full account of his research, sparing no detail, <u>which is permanently filed in a central repository</u>...[and] a two-page abridgment, which is published in journal form and numbered for identification, and a conventional abstract, which is reproduced on cards (border-punched preferably). A subscriber may receive the journal and the cards or, if he prefers, just the cards and may order by postcard photocopies or facsimile reproductions of the full papers <u>on deposit</u>. (J783).

Similarly, note the underlined constructions in the following abstract. Note that "<u>are described</u>", which can be regarded as the 'source' of "(system) <u>described</u>", is not deletable. This is also true of "<u>Store...</u> information" which precedes "<u>this stored</u> information" and of "<u>retrieval</u>" preceding "the <u>retrieved</u> bibliography". Note especially three very puzzling NP's : "<u>the punched card information</u>", which is relatable to the preceding sentence about 'abstracts being coded for reproduction on IBM cards'; "<u>an annual index to...</u>"and "<u>This book-form index-essentially the output of the Center's retrieval equipment</u>" are equivalent to "<u>The retrieved bibliography</u>" and the NP with "<u>cross-indexed bibliography</u>" to something like 'cross-indexed information'.

(32)

Methods used by the Thermophysical Properties Research Center for handling <u>collected bibliographical information</u> on thermophysical properties <u>are described</u>. Details are given regarding the method of coding abstracts for reproduction on IBM cards, particularly the coding of substances and the system of substance classification and notation used. Procedures used to <u>store the punched card information</u> on magnetic tape are described. Systems of <u>retrieval</u> of <u>this stored information</u> are delineated including a description of the organization of <u>the retrieved bibliography</u> and cf the machine techniques involved. <u>The system described</u> is claimed to be extensible to some other portions of scientific literature. Announced in this paper are the Center's plans to make available an annual index to the literature of thermophysical properties, expected to start publication in 1959. <u>This book-form-index--essentially the output of the Center's retrieval equipment</u>-- is described in prospect as a definitive, <u>cross-indexed bibliography</u> of every paper, book, report, or patent containing thermal property information. (J 985).

Also note that the above abstract displays the kind of relations between deixis and relative clauses discussed in the preceding section.

One last argument for regarding relative clauses as basic is related to what has been said about definitions, and about the occasional absence of lexical match between a relative clause and its discourse source as posited in the Annear-Langendoen hypothesis. The argument has to do with the relations between surface nouns and NP's with relative clauses (on this : Bach 1968). That a theory must somehow state the relatedness is clear from examples like the following:

(33)

(a) a machine that passively classifies its inputs (J 483).
(a') machines ... are only able to retrieve what has been put into them (J 364).

(a'') the machine is neither able to think about the information put into it nor able to forget obsolete information (J 364).
(b) the stored information or file of information (J 68).

Langendoen, who follows Bach 1968, attempts to go a step further by extending Bach's proposals in terms of his own proposals about relative clauses. Bach, let us say, had claimed that (34) (b) was the deep structure of (34) (a). Going a step further, Langendoen claims that this deep structure is actually (c) :

(34)

(a) An expert will speak.
(b) Someone who is an expert will speak
(c) Someone is an expert. He will speak.

(Langendoen 1970, 148)

My first objection to such a proposal is that it sounds absolutely ludicrous to posit a discourse source for various noun-phrases with adjectives and participles (which the transformational tradition equally oddly relates to relative clauses). Consider, for instance, expressions like "human being", "clerical person", "living object", and names of institutions like "American Council of Learned Societies".

My second objection is that I see no way of generating
definienda or expressions similar to definitions like "_the_
stored information OR file of information" (J 68) on the basis
of Langendoen's proposal.

My third objection is a more general one which, however,
is relatable to the first two. Following George Lakoff
(particularly 1969c, 3 and 29), I would say that language
offers the possibility of expressing the same information in
two mutually incompatible but equally basic ways, namely
by asserting this information (that is, presenting it as
something new) or, on the contrary, by presupposing that
it is known by the listener or reader; and that relative
clauses _and_ lexical items are precisely the most important
means by which _presupposed_ information is conv^ved. For
instance, George Lakoff would say that the sen nce, '_John's_
murderer reads dirty books'"pres·upposes" that John has been
murdered and that some books are dirty, and that all the
sentence asserts as something new is the 'reading' relation
between these two presuppositions (_loc. cit._).

If the arguments presented in this section are correct,
BOTH discourses and relative clauses are basic (i.e. should be
generated in the base), and the question is then _how_ relative
clauses are to be generated. In answer to this question, I can
only put forward the general suggestion that the generation
of relative clauses should be based on Langendoen's concept
of projection (1967). I shall not attempt to go beyond this
general suggestion, however, because of complications arising
with asymmetric conjunction. These are dealt with in the
next section, and I shall return to the problem in chapter 3.4.

3.3.3. Asymmetric conjunction and documentation.

This section is devoted to a preliminary exploration of
what underlies asymmetric 'and' in various expressions having
to do with the concept of documentation. I feel that this
concentration on examples from this conceptual field is
necessary to serve the requirements of distinctiveness of
semantic analysis as defined in Part I of this thesis. On
the other hand, I wish suggest that R. Lakoff's asymmetric
conjunction is an important mechanism in language and that
it is by no means an idiosyncracy of the conceptual field of
documentation.

I believe that some of the following sentences are
instances of asymmetric and, the written version of which may
be a comma or a comma + and . I shall not further discuss these
examples but wish to present them here in order to suggest
the relevance of the ensuing discussion to conceptual fields
other than documentation. With some qualifications to be
discussed below, one fairly clear criterion for regarding a
conjunction as asymmetric is that the order of the conjuncts
cannot be reversed without producing some oddity, or some
change in meaning. Compare the symmetric example in (35) (a)
with the asymmetric examples in (b), (c), and (d) and the same
examples with the order reversed:

(35)

(a) Summary of symposium on aeronautical documentation
    in the several NATO countries, including description
    of efforts in Canada, France, Holland, Federal
    Republic of Germany, U.S.A., and Italy. (J 31)
(a') ... efforts in Italy, U S.A., Federal Republic of
    Germany, Holland, France, and Canada.
(b) Selection, training, and relocation of personnel
    (J 999).
(b') ? Relocation, training and selection of personnel.
(c) Results of tests... are ... presented and analyzed
    (J 1111).

(c') ? Results are analyzed _and_ presented .
(d) Experience in setting up/installing _and_ using the...
    system (J 122).
(d') ? using _and_ setting up the system.

Before turning to examples from the field of documentation,
let me further examine the effect of reversing the order of
conjoined sentences in asymmetric (36) and symmetric conjunc-
tion (37).(36) (a) is adapted from one of Jespersen's
examples to which I shall again return below; (36) (b) shows
that reversing the order of the conjuncts changes the meaning
of (a). This, however, is not the case with symmetric conjunc-
tion, as is shown in (37) (a) and (b).

(36)

(a) The doctor arrived rapidly AND the patient soon
    recovered.
(b) The patient soon recovered AND the doctor arrived
    rapidly.

(37)

(a) John works for IBM AND he plays tennis very well.
(b) John plays tennis very well AND he works for IBM.

Now, reversing the order of the conjuncts is by no means
the only test for establishing the asymmetric nature of
conjunction.(38) (a) and (b) are borrowed from Jespersen
(1965, 137 and 1966, 316). The remaining sentences of (38)
are clearly related to the first two in meaning, and this,
I suggest must be regarded as evidence for the asymmetric
character of the conjunctions ('_and_' or semi-colon) in (a).

(38)

(a) The Doctor arrived extremely quickly AND examined
    the patient uncommonly carefully; she recovered
    very speedily.
(b) The Doctor arrived extremely quickly AND examined
    the patient uncommonly carefully; THE RESULT WAS THAT
    she recovered very speedily.
(c) The doctor's extremely quick arrival AND uncommonly
    careful examination of the patient BROUGHT ABOUT
    her very speedy recovery.

(d) BECAUSE the doctor arrived quickly AND (BECAUSE he) examined the patient very carefully, she soon recovered.

(e) THANKS TO the doctor's quick arrival and (THANKS TO his) careful examination of the patient, she soon recovered.

(f) The doctor arrived quickly AND examined the patient very carefully IN ORDER THAT she might soon recover.

(g) The doctor arrived quickly AND examined the patient very carefully SO THAT she soon recovered.

(h) The patient WHO had been examined very carefully by the doctor WHO had arrived very quickly recovered very speedily.

(i) IF the doctor arrives very quickly AND examines the patient carefully, she will soon recover.

(j) IF the patient is to recover soon, the doctor must arrive very quickly and examine her very carefully.

(k) The patient will not recover UNLESS the doctor arrives very quickly and examines her very carefully.

(l) The doctor's quick arrival AND careful examination of the patient will ENABLE her/will MAKE IT POSSIBLE for her/IS NECESSARY for her to recover speedily.

(m) BY arriving very quickly AND (BY) examining the patient very carefully, the doctor ENABLED her TO recover speedily.

(n) The patient's speedy recovery DEPENDS ON the doctor's quick arrival AND on his careful examination.

The examples given below in (39) can be regarded as indirect evidence for the asymmetric analysis of (38) (a). The sentences given in (39) (b)-(k) force us to interpret (39) (a) as an asymmetric conjunction in that they force us to interpret the 'and' as some "causal or implicational link" (R. Lakoff). The only type of construction that does not seem to produce a change of meaning is relative clause (391,m). In other words, one may say that when the meanings of the conjoined sentences do not impose an asymmetric interpretation, a conjunction of sentences and the corresponding sentence with a relative clause are equally ambiguous between the symmetric and the asymmetric reading.

(39)

    (a) John works at IBM and he plays tennis very well.
    (b) John's work at IBM RESULTS in his playing tennis very well.
    (c) BECAUSE John works at IBM, he plays tennis very well.
    (d) THANKS TO his work at IBM, John plays tennis very well.
    (e) John works at IBM IN ORDER TO be good at tennis.
    (f) John works at IBM SO THAT he plays tennis very well.
    (g) IF John works at IBM, he will be good at tennis.
    (h) IF John is to play tennis well, he will have to work at IBM.
    (i) John will not be good at tennis UNLESS he works at IBM.
    (j) John's work at IBM ENABLED/IS NECESSARY FOR him to play tennis well.
    (k) BY working at IBM, John is ABLE to be good at tennis.
    (l) John, WHO works at IBM, is good at tennis.
    (m) John, WHO is good at tennis, works at IBM.

Let us now return to the field of documentation. The importance of conjunction in this field already appears in Webster's definition of 'documentation' :

(40)

The assembling, coding and dissemination of recorded knowledge comprehensively treated as an integral procedure utilizing semantics, psychological and mechanical aids, and techniques of reproduction including microcopy for giving documentary information maximum accessibility and usability.

Applying the reversibility test we find that "the assembling, coding and dissemination of recorded knowledge" is an instance of asymmetric conjunction : for instance, "dissemination, coding, and assembling of recorded knowledge" is either unacceptable or different in meaning from the passage of Webster's definition.

However, the reversibility test doesn't always work.
In fact, there are cases in which conjunction seems to have
the effect of scrambling the expected order of words :

(41)

f. documentation system is defined as a complex pattern
of interacting and interdependent processes which
facilitates the use of recorded knowledge through its
<u>presentation</u>, <u>reproduction</u>, <u>publication</u>, <u>dissemination</u>,
<u>acquisition</u>, <u>characterization</u>, <u>storage</u>, and <u>retrieval</u>
(J 107).

Because the reversibility test does not always work, it
seems that a better procedure for defining asymmetric
conjunction (as well as establishing that one has to do
with it in a particular sentence) is the one adopted above
with Jespersen's example : it consists of showing that
asymmetric conjunction is relatable to various other types
of constructions.

To illustrate the procedure with another brief example,
I suggest that an expression like 'history and development'
(J 532) needs to be understood in terms of the following,
parallel or relatable expressions : <u>historical development</u>
(J 25), <u>history and further development</u> (J 46), <u>history of their</u>
<u>development</u> (J 486). Similarly, one can examine the correlation
between "<u>the design, operation, and maintenance of the weapons</u>
<u>of World War II</u>" (J 828) and, say, <u>the maintenance of weapons</u>
<u>designed and operated in World War II</u>. In what follows I shall
apply the same technique and present   attested examples
which have to do with documentation.

In the Janaske corpus, I have found a number of construc-
tions that are relatable to asymmetric conjunction. In (42)
(i)-(vi), I present the main types of these constructions
in a nutshell, as the constructions that are relatable to the
conjunction "the assembling (Vling), coding (V2ing), and
dissemination (V3ing) of knowledge (NP)"(Stockwell et. al.
1968 would write N instead of V, but this is irrelevant to
my present purposes)

<center>(42)</center>

(i) The TASK of Vling AND V2ing AND V3ing NP.
(ii) NP is Vled (in order) TO be V2ed AND V3ed.
(iii) The V2ing AND V3ing of NP BY Vling it.
(iv) The V3ing of NP (which has been previously) Vled
   AND V2ed.
(v) The Vling of NP RESULTS IN the V2ing and V3ing of NP.
(vi) (The) Vling AND (THEN) V2ing AND (THEN) V3ing of
   NP.

(i) The examples quoted below in (43) (a) and (b) show
   that when there is a series of verbs such as that represented
   above as V1, V2, and V3, these verbs or cognate nominals
   can appear in an NP whose head is a noun such as task, or
   operation, with this noun in the SINGULAR. The singular
   appears to be one of the marks by means of which the
   processes denoted by the verbs are conceived of as a whole,
   as something global. Other marks having the same value as
   the singular are adjectives such as 'total'. The singular
   also appears with the word system as head, as is shown
   in (43) (c), which can be compared with the part of
   Webster's definition repeated in (d) and with the equally
   complex example of (e).

<center>(43)</center>

(a) The task of processing and retrieving information
   (J 66).
(b) The total storage and retrieval operation (J 970).
(c) a storage-retrieval system (J 978).
(d) The assembling, coding and dissemination of recorded
   knowledge comprehensively treated as an integral
   procedure...

(e) Documentation : <u>Complete cycle</u> of information service.
"Documentation is best differentiated from normal library service by the extent to which it is concerned with <u>a complete-cycle service</u> providing information. This cycle involves the identification, recording, organization, storage, recall, conversion into more useful form, synthesis, and dissemination of the intellectual content of print and other recorded materials." (J 890).

(ii) The correlation between asymmetric <u>and</u> and expressions of purpose is well-evidenced in the Janaske corpus. Compare, for instance, "<u>develop</u> AND <u>apply</u> automatic techniques" (J 819) and "the system was <u>developed</u> FOR <u>use</u>" (J 115). The other examples, which are given in (44), have to do with the concept of documentation. The example with "<u>so that</u>... <u>can</u>" (b') is treated as an expression of purpose (rather than result) because it can be paraphrased by '<u>in order to (that)</u>'.

(44)

(a) Abstracting, <u>coding</u>, AND <u>searching</u> the metallurgical literature (J 508).

(a') the index of Chemical Abstracts could be <u>coded</u> mechanically on a computer FOR LATER <u>searching</u> on a computer (J 239).

(b) storage (= filing) AND retrieval

(b') Most mechanized information retrieval systems depend upon <u>extracting</u> information from documents, <u>processing</u> such information in some manner AND FINALLY <u>placing</u> this information <u>in a file</u> SO THAT the information CAN BE <u>retrieved</u> by machine (J 321).

(c) literature <u>searching</u> AND <u>retrieval</u> (J 1025).

(c') multiple aspect <u>searching</u> FOR information <u>retrieval</u> (J998).

(c'') The essential PURPOSE of literature <u>searching</u> is TO <u>locate</u> those documents... (J 773).

(d) Subject matter <u>analysis</u> AND <u>coding</u> (J 769).

(d') the <u>analysis</u> of this information FOR <u>indexing</u> and <u>cataloging</u> (J 68).

(iii) One further construction which is relatable to asymmetric
conjunction is given below in (45) (b) :

(45)
(a) The essential PURPOSE of literature <u>searching</u> it
TO <u>locate</u> those documents... (J 773).
(b) papers <u>located</u> BY the direct <u>search</u> of abstracting
journals (J 156).
(c) <u>search</u> AND <u>locate</u>

A comparison of (45) (a) and (b) suggests that asymmetric
conjunction of the type under investigation is relatable
to expressions of purpose as well as to converse
<u>instrumental</u> expressions. Further evidence for the cor-
relation between purpose and instrumental expressions can
be found in the following, more complex, examples :

(46)
(a) The assembling, coding and dissemination of recorded
knowledge comprehensively treated as an integral
procedure (...) FOR giving documentary information
maximum accessibility and usability (from Webster's
definition of <u>documentation</u>).
(b) (a documentation system...) facilitates the use
of recorded specialized knowledge THROUGH its
(...) storage and retrieval (J 107).

(iv) Another type of construction related to asymmetric
conjunction is illustrated in (47) (a) and (b), namely
constructions relatable to relative clauses with or
without indications of time sequence. (c) illustrates
the possibility of such constructions being replaced
by an NP with a head noun corresponding to the <u>Ved</u>
form ('<u>stored information</u>'/'<u>information STORE</u>').

(47)
(a) <u>storage</u> AND <u>retrieval</u> of information (many examples)
(a') methods for <u>retrieving</u> <u>storED information</u> (J 174)
(a'') retrieving <u>information WHICH is or has been
previously stored</u>.
(b) Semi-automatic <u>indexing</u> AND <u>encoding</u> (J 321).
(b') <u>Encoding</u> of information PREVIOUSLY <u>indexED</u> (J 511).
(c) a <u>storage</u> AND <u>search</u> theory (J 250).
(c') <u>search</u> OF an information STORE (J 943).

Also compare (48) (a) with the _Ved_ forms underlined
in (48) (b), an abstract already quoted in 3.3.2. :

(48)

(a) (Center established for the purpose of...)
by _collecting, processing, storing,_ and _retrieving_
scientific information (J 1).

(b) Svstems and orocedures developed for the search,
coding ana mennenizea processing oi bibliographic
information on thermophysical properties.
Methods used by the Thermophysical Properties
Research Center for handling _collected bibliogra-_
phical information on thermophysical properties
are described. Details are given regarding the
method of coding abstracts for reproduction on
IBM cards, particularly the coding of substances
and the system of substance classification and
notation used. Procedures used to store the
punched card information on magnetic tape are
described. Systems of retrieval of _this stored_
information are delineated including a description
of the organization of _the retrieved bibliography_
and of the machine techniques involved. The
system described is claimed to be extensible to
some other portions of scientific liteeature.
Announced in this paper are the Center's plans to
make available an annual index to the literature
of thermophysical propeвties, expected to start
publication in 1959. This book-form index --
essentially the output of the Center's retrieval
equipment -- is described in prospect as a
definitive, _cross-indexed_ bibliography of every
paper, book, reоort, or patent containing thermal
property inforiu.tion. (J 985).

(v) Yet another type of construction which is relatable to
asymmetric conjunction is illustrated in (49). Compare
(a) with (b), (c) and (d) :

(49)

(a) Proposes a system providing for a conjoint _search_
of separate but correlated record stores AND
an uрсional _printout_... (J 473).

(b) Successful _searches_ by this mechanized system
RESULT IN the _printout_... (J 517).

(c) _The terminology and coding_ that RESULTED IN
selection of some abstracts not obviously
pertinent (J 821).

(d) (services on the part of the library) involve
    a regu' scrutiny of ... periodicals with [= AND]
    CONSEQUENT indexing and distribution of the
    RESULTING information (J 446).

The following example illustrates an even more complex
situation, where the expression relatable to result is
embedded in a manner NP ('in an organized manner'):

(50)

Storage of this information in an organized manner
that PERMITS [= RESULTS IN] rapid retrieval (J 494).

(vi) Next to instrumental, purpose, result and consequence
     expressions another type of construction relatable to
     asymmetric 'and' can be called time sequence. It has
     been examined by Robin Lakoff in the form 'and then',
     but this form is by no means the only one. Besides, and-
     conjunction, is not the only construction in which
     indications of time sequence can occur, as already
     appeared from some examples quoted above and repeated
     below. The time factor can also be as it were factored
     out and made into an NP, as is shown in (51) (f) and (g).

(51)

(a) Encoding of information PREVIOUSLY classified (J 651).
(b) Most mechanized information retrieval systems
    depend upon extracting information from documents,
    processing such information in some manner, AND
    FINALLY placing this information in a file so
    that the information can be retrieved by machine
    (J 321).
(c) The index of Chemical Abstracts could be coded
    mechanically on a computer FOR LATER searching
    on a computer (J 239).
(d) (possible) for the abstracts THEN to be automatical-
    ly encoded FOR SUBSEQUENT searching (J 656).
(e) (documentation is defined as) a technique for
    communication with unknown persons IN THE INDEFINITE
    FUTURE (...) the flow of information to the unknown
    user IN THE FUTURE (J 81).

(f) the TIME gap between the <u>generating</u> of information <u>and</u> its availability (J 350).

(g) <u>Shortening</u> the TIME lag between original discovery_{and} engineering <u>application</u> (J 987).

In the next chapter (3.4.), I shall return to the problem of asymmetric conjunction. Until then I shall stick to the notation introduced above in (42).

3.3.4. Lexical presupposition and Lyons's implication.

About her examples of asymmetric conjunction, Robin
Lakoff claims that

> Taken out of context, the conjuncts would not retain
> this causal or implicational link, so that part of the
> meaning of the sentence would be lost - the idea that
> each one led up to the next (Robin Lakoff 1969, 23).

As pointed out in 3.1.4., I suggest that this is only partly
true, and I wish to argue in this section that asymmetric
conjunction in the field of documentation displays two
types of presuppositions. One is the type of lexical
presupposition which has been mostly discussed by Fillmore.
In this type, the presuppositions are statable as part of
the properties of lexical items, such as verbs, and the
presuppositions continue to hold when a sentence with the
verb in question as main verb is used in isolation. Fillmore's
analysis of 'borrow' is a good example : with or without
context, you cannot help associating the presupposition (C had
B before time T, with a sentence 'A borrowed B · from C at
time T' (Fillmore 1968a, 52). Similarly, it is a property of a
predicate like 'retrieve' that you can only 'retrieve' s· thing
which is lost, hidden, which you do not possess, etc.
Or take the pair 'input-output'. I believe that it is a property
of the WORD 'output' that it presupposes 'input'. I give this
example because a sentence of the Janaske corpus which is
relatable to 'input-output' happens to be the only place at
which the verb 'presuppose' occurs in the whole corpus :
"To get relevant and valid information out of a machine
presupposes that it has been put in with a most exacting
pattern of forethought" (J 179) (cp. 'throughput' which, I
gather, has crept into the computer jargon).

Robin Lakoff's point is only applicable to what can be
called non-lexical or situational presupposition, which
can be illustrated as follows. Imagine a situation in
which John usually performs the following actions one after
another : he gets up, then washes, dresses, has breakfast,
and goes to school. If John is in bed, his mother can call
him by saying "John, it's time to—" + ANY of the above
sentences and any of these can be said to presuppose all the
others. But such a presupposition cannot be regarded as a
property of the verbs in question : people can have breakfast
in bed, at school, at noon, etc.

In the present section, I shall represent the conceptual
field of documentation in a fashion similar to that adopted
in the preceding section when "the assembling, coding, and
dissemination of knowledge" was symbolized as "V1ing AND
V2ing AND V3ing of NP". On this basis I propose to define
V1ast as the last V in a given sequence of V's that are
relatable to each other by asymmetric conjunction. Thus
"dissemination" is V1ast in the above example. If 'use' is
added as V4 to the above ("the assembling, coding, dissemina-
tion and use of knowledge"), it becomes V1ast. Furthermore
a situation in which, say, a V1 presupposes a V2 will be
described as 'forward presupposition' and the converse situa-
tion as 'backward presupposition'. One everyday example of
forward presupposition is the use of GO OUT in the meaning
of "mix in society; attend social functions, etc. : She
still goes out a great deal, even at seventy-five" (Hornby).
In all this discussion, then, I shall go on speaking — as
Fillmore has often done — of 'presupposition of verbs' as a
convenient abbreviation for 'presupposed belief conditions
for an appropriate use of sentences containing these verbs as
main verbs - or cognate nominal constructions with the cognate
nominals acting like main verbs in a sentence' (cp. Fillmore
1970b).

My main purpose in proposing such concepts as <u>Vlast</u>
is that of explaining the use of <u>single</u> <u>V</u>'s which, as far
as their meaning is concerned, appear as abbreviations of
string of conjoined V's. Compare, for instance, the frequent
phrase "<u>information storage and retrieval</u>" and the abbrevia-
tion to "<u>retrieval</u>" in "On <u>retrieval</u> system theory" (J 1044).
I shall discuss successively :

    (i) presuppositions associated with <u>Vlast</u>;
    (ii) forward and backward presuppositions;
    (iii) I shall then turn to problems associated with
          Lyons's <u>implication.</u>

A more formal treatment of some of these questions will be
presented in the next chapter (3.4.).

(1) One example of <u>Vlast</u> is, I believe, the word '<u>documentation</u>',
and the cognate verb '<u>document</u>', one of the definitions of
which is "to supply with documents" (Webster's). From
the following abstract, for instance, I suggest that it is
possible to reconstruct an asymmetric conjunction which
would include among others "<u>cover</u>" ("literature"),
'<u>process</u>' (with "mechanical aids"), '<u>search</u>' (on the basis
of "searcher"), and "<u>document</u>". Yet only <u>documentation</u>
(in the title) and "<u>document-</u>" (in sentence one) are
mentioned at the beginning. What they presuppose, however,
is ASSERTED by the rest of the text. A similar example
with '<u>dissemination</u>' is given in (53).

(52)

    (a) Problems of <u>documentation</u> of the literature of
        atomic energy.
        The problems of <u>documenting</u> the literature of
        atomic energy are many. The volume of literature
        to be <u>covered</u> is large, it draws on a wide range
        of subjects, security difficulties place barriers
        against the <u>searcher</u> and cause duplication of
        effort. <u>Mechanical aids</u> are desirable. The status
        of the documentalist should be improved, since
        it is difficult to draw the line between his
        responsibilities and that of the scientist (J 371).

(b) <u>Proposed analysis</u>. "Docume:'ng the literature of
atomic energy" presupposes . it precedes in the
conjunction "cover, process, arch and document
the literature of atomic energy", and it can be
regarded as an abbreviation of this conjunction.

(53)

(a) The <u>dissemination</u> of the knowledge of Soviet
scientific work in Western countries.
Paper devoted to discussion of sources of Russian
scientific information as well as the means of ob-
taining sources specific to each. Attention is given
to available as well as needed areas of <u>abstracting</u>
and <u>indexing</u> services. The "language barrier" and
consequent role of <u>translation</u> of Soviet periodicals
together with current enterprise in this field is
discussed. Data are presented in summary form on
the proportion of the published scientific literature
from the Soviet Union currently available in England.
(J 945).

(b) <u>Proposed analysis</u>. "The <u>dissemination</u> of..."
presupposes what precedes in the conjunction
"abstracting,indexing, translation, and dissemina-
tion of..." (or possibly : "translation, abstracting,
...") and it can be regarded as an abbreviation of
this conjunction.

By similar reasoning, I suggest that one can account
for an expression '<u>documentation center</u>'. A documentation
center does not merely 'supply' people with documents.
One is also tempted to account for expressions like
'<u>information center</u>' or '<u>information system</u>' in this
fashion (an alternative explanation is presented below).
Evidence for this is the following attested (and, to me,
slightly odd) sentence in which '<u>communicate</u>' seems to
function as <u>V1</u> and '<u>inform</u>' as <u>V2</u> : "The objective
of scientific <u>communication</u> is to <u>inform</u>" (J 247; compare:
'communicate and inform', not attested).

Other instances of _Vlast_ are underlined in the following
sentences. Some are conceptually equivalent to 'documenta-
tinn' and some are actually definitions of the concept:

(54)

(a) _Finding_ what's known. Information retrieval techniques
to make the vast store of knowledge conveniently
accessible... the intellectual job of matching
what's known and what's wanted (J 407).

(b) Documentation is defined as a technique for
communication ... (J 81).

(c) Scientific documentation -- the operation of
bringing the facts elucidated by scientific
research from those who produce them to those who
can make use of them -- (J 97).

(d) The problem of _bridging_ the gap between the
individual who wants the information and the
records containing the information (J 205).

(e) The basic problem is to _increase the mental contact_
between the reader and the information stored (J 474)

(f) _Keeping_ research _in contact with_ the literature

(g) The man-document _communication_ problem (J 1110).

(h) The problem of _getting_ information _into the hands_
_of_ the scientist who needs it (J 418).

(11) Let us now consider forward and backward presuppositions.
In order to understand the examples presented below, one
must bear in mind that punched-cards in documentation
are strictly speaking a _storage_ medium. In other words,
in a conjunction such as "_collect_ (V1), _abstract_ (V2),
code (V3), _store_ (V4), _retrieve_ (V5) and _disseminate_ (V6)".
punched-cards would normally be associated with V4 as in,
say, 'stored on punched-cards'. This, however, is not
necessarily so. In the following examples, '_punched-_
cards' are associated with '_coding_' (V3), '_collecting_'
(V1), and '_abstracting_' (V3), which I would regard as
_V_'s that are PRESUPPOSED by 'stored on punched-cards'.
Similarly, given the conjunction "collect (V1) and handle
(V2) information", note the shift in the same abstract from
V2 to V1 exemplified in (55) (c) :

(55)

(a) Coding of Hammett RHO-values on punched cards.
McBee Keysort cards are used to collect data con-
taining application of the quantitative relations
between the structure and the reactivity of organic
compounds by the Hammett equation. (J 469).

(b) The use of four-hole randomly punched cards
for abstracting publications and reports
into IBM cards.
Coding of abstract information on IBM cards is limit-
ed by the 80-column capacity of the cards. A ran-
dom code of two, three, or four holes successively
increases the capacity to code categories of informa-
tion and still retains space on the card for coding
other identifying information on the referenced pub-
lication or report. The code, storage, and retrieval
processes are described. (J 904).

(c) Information handling in a large information system.
Concerns research into the collection of information
on chemical compounds found in edible plants.
The study was conducted at the Low Temperature
Station for Research in Biochemistry and Biophysics,
University of Cambridge, England. This paper is
a description of the resulting system, and a report
of various problems encountered in its development
(J 167).

Lastly, forward and backward presupposition account for
the possibility of having expressions consisting of V's
that are non-contiguous in the asymmetric conjunction
by means of which I propose to represent the concept of
documentation. The expressions have the same global
meaning as 'documentation' by virtue of forward and
backward presupposition. Thus, "organization and dissemina-
tion" in the following example is equivalent to 'documenta-
tion' and can be said to stand for something like
'collection, organization, storage, retrieval, and
dissemination'; I shall attempt to clarify what I mean
by 'something like' at the end of this section in the
light of what Lyons calls implication :

(56)

It is concluded that the organization and dissemination
of scientific information is a professional activity...
(J 967).

(111) In the last published formulation with which I am
acquainted, Lyons illustrates his proposals about _implica-_
tion rules with the sentence 'I bought some flowers'.
He suggests that this sentence _implies_ a "disjunction"
or a conjunction (actually, he speaks of 'co-ordination'
- between inverted commas) of sentences with hyponyms
of 'some flowers'. This can be written as : 'I bought
some flowers' IMPLIES 'I bought some tulips' AND/OR 'I
bought some roses', etc. (Lyons 1968, 455). Just a few
pages before the passage just referred to, he defines
implication in terms of "_implicit_ assertion and denial"
(_op. cit._, 445). This, it seems to me, is a more fortunate
characterization than the other formulation which I have
illustrated by means of AND/OR. I suggest that in the
above example the AND can be equated with "(implicit)
assertion", and the OR with "(implicit) _denial_", and that
this OR can be replaced by AND NOT. Sticking to the kind
of informal representation adopted by Lyons (a more formal
symbolisation would require more knowledge of negation
than is, I think, currently available), the above example
can be written as in (57) :

<div align="center">(57)</div>

I bought some flowers IMPLIES
AND [± NOT (I bought some tulips)]
AND [± NOT (I bought some roses)]
etc.

Obviously, at least _one_ ± NOT in (57) must be -NOT.
Moreover, the implication is _bilateral_ in the above example.
Lastly, note that the formulation of (57) captures the
indeterminacy of implication in the same way as does the
formulation with AND/OR. (on the reasons for positing
AND in the first conjunct and an _and_-deletion rule, see
Lakoff and Peters 1969, 114 fn. 2). On the basis of Leech
1969 (35-38), similar rules can be written for an implica-
tion relation between sentences in which the superordinate

term and the corresponding hyponyms are predicates. This
is what we need here, for sentences belonging to the
conceptual field of documentation. Provisionally, I ignore
the distinction - as does Lyons - between symmetric and
asymmetric 'and'. Now consider the following translation of
the main headings of the Gardin-Lévy concordances (cp. an
almost identical translation given in Appendix 2) :

(58)

1. Scientific and technical information : generalities.
2. Professional problems.
3. Information science.
4. Related sciences.
5. Networks of scientific and technical information.
6. Scientific and technical information institutions.
7. Functioning of scientific and technical information
   networks.
8. Sources of information.
9. Techniques of reproduction.
10. Information handling : generalities.
11. Collection of information.
12. Documentary analysis : generalities.
13. Cataloguing.
14. Abstracting.
15. Synthesis.
16. Classification and indexing.
17. Extracting, tabulation and index production.
18. Translating.
19. Coding.
20. Storing.
21. Searching.
22. Dissemination.
23. Machines and equipment.
24. Terminology.
25. Standards.

Note that the headings contained in the above boxes
can be conjoined. One can say, for instance :

(59)

(a) information <u>handling</u> consists of collecting,
cataloging, abstracting, synthesizing, classifying,
and indexing, extracting, translating, coding,
storing, searching, and disseminating scientific
information.

(b) documentary <u>analysis</u> consists of cataloguing,
abstracting, synthesizing, classifying and
indexing, extracting, translating, coding and
storing scientific information.

(59) (a) and (b) can be rewritten as implication rules
as in (60) :

(60)

(a) <u>handling information</u> IMPLIES

AND [± NOT (collecting information)]
AND [± NOT (cataloguing information)]
AND [± NOT (abstracting information)]
AND [± NOT (synthesizing information)]
AND [± NOT (indexing information)]
AND [± NOT (extracting information)]
AND [± NOT (translating information)]
AND [± NOT (coding information)]
AND [± NOT (storing information)]
AND [± NOT (searching information)]
AND [± NOT (disseminating information)]

(b) <u>analyzing information</u> IMPLIES

AND [± NOT (cataloguing information)]
AND [± NOT (abstracting information)]
AND [± NOT (synthesizing information)]
AND [± NOT (indexing information)]
AND [± NOT (extracting information)]
AND [± NOT (translating information)]
AND [± NOT (coding information)]
AND [± NOT (storing information)]

(c) = (a), with the replacement of the appropriate
items by 'analyzing information' (.b) :
<u>handling information</u> IMPLIES

AND [± NOT (collecting information)]
AND [± NOT (analyzing information)]
AND [± NOT (searching information)]
AND [± NOT (disseminating information)]

The implication relation in (60) (a) and (b) can be regarded as bilateral. Now, the ± NOT's in the above rules are not quite correct. First, like in the example borrowed from Lyons, there is an obvious _logical_ reason for imposing that at least _one_ NOT should be marked minus. Second, and also for _logical_ reasons, the 1st sentence in (60) (a) must be -NOT: '_Collecting information_' must be -NOT (i.e. assertive), because you must '_collect information_' in order to be able to '_disseminate information_'. This would be statable in case theory by saying that some institution must be the Goal of 'collecting information' in order to be the Source of 'disseminating information'. Third, there are _pragmatic_ reasons which only the expert of the field could state for deciding that certain operations of documentation are necessary, while others are not. Since the Gardin-Lévy concordance is not specific about this, I shall just assume in what follows that '_abstracting_', '_synthesis_', '_translating_', and '_coding_' are optional, and that the other operations are necessary. Also for _pragmatic_ reasons, an operation such as '_translating_' may be necessary : for instance, in the abstract quoted above in (53), about "the dissemination of the knowledge of _Soviet_ scientific work in _Western_ countries" (J 945).

I suggest that _lexical_ presupposition (as in Fillmore's 'borrow' sentence) corresponds to what I call logically - based implication, while the other type of presupposition (e.g. the relation between "the police arrived" and "everybody swallowed their cigarettes" in Robin Lakoff's example) would correspond to my pragmatically-based implication. Ignoring the uncertainties just noted, implication rules of the type given in (60) for the _metalanguage_ can be regarded as necessarily 'right' or 'correct'.

But one <u>cannot</u> say that they are <u>true</u>. On the basis of a
recent discussion by Bierwisch, we can say that implica-
tion rules for the metalanguage cannot be assigned a
truth-value because they are <u>stipulative definitions</u> :

> A stipulative definition stipulate that, whatever
> the word may mean in other communications or even in
> earlier parts of this communication, it is for the
> rest of this communication to be taken as having
> <u>no meaning whatever</u> except the one now stipulated. Any
> previous meanings are thereby abolished for the remainder
> of this communication (Robison 1950 quoted in Bierwisch
> 1969, 77).

In the rest of this section, I wish to mention three
further points, which concern the relations between object-
language and metalanguage in this thesis. First, the headings
of the metalanguage are <u>by definition</u> NOT subject to presup-
position as are the corresponding verbs in the object-language.
This is also due to the stipulative character of the meta-
language. To take just one example, a heading like DISSEMINA-
TION (Nr. 22 in the Gardin-Lévy concordance) is, by definition,
understood as 'dissemination' in the strict or narrow sense.
It is by virtue of this that the metalanguage can provide
the semantic PRIMES for the present investigation. As will be
suggested, by my third point below, I cannot think of another
way of defining semantic primes for this study.

My second point has to do with hesitations which I have
experienced on several occasions when indexing abstracts
of the Janaske bibliography. These hesitations can be explained
by returning to an everyday example with to GO OUT. Suppose
a teenager tells his parents : '<u>I am going out</u>'. This is
ambiguous, depending on whether or not you add various
presuppositions (such as '<u>to attend a party</u>') to the meaning
of, say, 'leaving the house'. Similarly, in the conceptual
field of documentation, a <u>V</u>last such as, say, '<u>dissemination</u>'
always presupposes (at least some of) the other <u>V</u>'s.

Thus I would say that the first abstract quoted below
(in 61a) presents such a problem. While it is clear that
its central topic is DISSEMINATING ("presenting", "disseminat-
ing") at least one presupposition shows up in the text,
namely in "classification number". In the abstract of (61)
(b), on the other hand, DISSEMINATING can be regarded as the
only topic. This is no doubt in part due to the fact that
no mention is made of processes that 'precede' dissemination,
and to the fact that the Vlast is not "disseminate" but
"provide" : "disseminating ... literature to the researcher...
provide the user with ..."
                                (61)

(a) New methods for presenting bibliographical
    information in the U.S.S.R.
    Two methods of disseminating scientific technical
    articles are described as used by the State Public
    Scientific and Technical Library attached to the
    Siberian Department of the USSR Academy of Sciences.
    The first method, which is suitable only for individ-
    ually prepared bibliographies or other cases
    requiring only a very few copies, involves the use
    of pockets the size and shape of catalog cards into
    which microfilm copies of articles are inserted. The
    bibliographic description, classification numoer and
    card number are printed on one side of the pocket,
    and the serial and classification numbers are shown
    on the first frame of microfilm. Production costs are
    about two rubles a card. For current information on
    a large scale, microcards are used. Since the av-
    erage article is only four pages in length, a modified
    type of card is prepared which reduces the original
    only five or six times. The cards can, therefore, be
    read with very simple magnifying glasses. Biblio-
    graphic information is photomechanically printed on
    a catalog card, on the back of which is pasted the
    photographic reproduction of the original. At the
    beginning of 1960 the Library opened a subscription
    list and inaugurated a pilot service of this kind for
    mechanical engineering and plastics. (J 551).

(b) Sassif (Self adjusting system of scientific
    information flow).
    Describes a system of disseminating newly published
    scientific literature to the researcher. This system
    will provide the user with (1) originals of all

> articles clearly of interest to him, (2) summaries
> of articles not in the scope of marginal interest
> to him, (3) summaries of articles not in the scope
> of the user's present range of interests to titillate
> his interests and relieve him of the burden of scan-
> ning a large body of literature of marginal interest.
> (J 141).

In the cases where, as in (61 a), 'dissemination' has or may
have a global meaning, I would suggest that it should be ge-
nerated by implication rules of the type illustrated in (60) from
an OBJECT-LANGUAGE sentence with a superordinate verb (e.g.
'information handling'). In the other case (61b), the under-
lying sentence would have as its verb 'disseminate' and rules
of presupposition would take over to state that, say, what
has the meaning of "dissemination" proper presupposes the
processes that precede in the field of documentation.

My third and last point is the following. One of the
major difficulties in dealing with the vocabulary of the
Janaske corpus (i.e. the object-language under investigation)
is that superficially different lexical items can correspond
to the same heading of the metalanguage. Different lexical
items with the same meaning commonly occur in the same
abstract : for instance, "information filing and retrieval =
arranging and finding" (J 60), "presenting = disseminating
information" (J 551), "data banking = information storage and
retrieval" (J 334). A generative semanticist might argue that
lexical decomposition would be able to display, and thus
predict, such equivalences between superficially different
predicates. Until the point can be demonstrated, one reason
for doubting that such a solution is achievable is that even
with such superordinate predicates as "handling", it may be
necessary to resort to a stipulative (i.e. linguistically
unmotivated) definition. An example is given below, in (62)
(a). In spite of the occurrence of "handling" in the abstract,
I believe that this text should be indexed by the terms

ANALYSIS (Nr. 12) and SEARCHING (Nr. 21) of the Gardin-Lévy
concordance rather than by HANDLING (Nr. 10). In (62) (b),
I have tabulated the object-language expressions and the
corresponding headings of the classification. Implication rules
between the various sentences tabulated are statable on the
basis of the fact (to be discussed in 3.5.) that in many
cases 'using a method (system, machine, etc.) to do something'
is equivalent to 'doing something with the method (system,etc.
in question)'.

<div align="center">(62)</div>

(a) <u>Methods used</u> by the Thermophysical Properties
Research Center <u>for handling collected bibliographical
information</u> on thermophysical properties are de-
scribed. Details are given regarding <u>the method of
coding abstracts</u> for reproduction on IBM cards, par-
ticularly the coding of substances and <u>the system of
substance classification and notation</u> used. <u>Proce-
dures used to store the punched card information</u> on
magnetic tape are described. <u>Systems of retrieval of
this stored information</u> are delineated including a de-
scription of the organization of the retrieved bibliog-
raphy and of the machine techniques involved. The
system described is claimed to be extensible to some
other portions of scientific literature. Announced in
this paper are the Center's plans to make available
an annual index to the literature of thermophysical
properties, expected to start publication in 1959.
This book-form index -- essentially the output of the
Center's retrieval equipment -- is described in pros-
pect as a definitive, cross-indexed bibliography of
every paper, book, report, or patent containing ther-
mal property information. (J 985).

(b)

	Object-language	Metalanguage
a	<u>handling information</u> which implies b - d	<u>ANALYSIS</u>, which implies b - c
b	coding classification notation	CODING ? CLASSIFICATION CODING
c	store	STORING
d	retrieval	(SEARCHING : see table 63 , below)

I now wish to end this section by presenting a table which gives some of the verbs, verbal expressions and cognate nominals found in the Janaske corpus that correspond to the headings of the Gardin-Lévy classification. Whenever it was possible, the order in which they are listed is the same as that adopted by Gardin-Lévy for their main entries ("groups") and these entries are given between brackets and capitalized. Some co-constituents of the words listed are given when necessary, for instance for some technical terms. A ± sign is added before a <u>V</u> of the Gardin-Lévy classification when I assume that the process or operation in question is optional. As will be noted, the same object-language word can appear under various headings. Also note that various lists have no corresponding descriptor in the "groups" of the Gardin-Lévy concordance which have been considered in this section. The typical Object of all the <u>V</u>'s listed is <u>information</u>, though it is by no means the only one. The typical Agent of <u>V1</u> to <u>V12</u> is some kind of institution, or some officer.

<div align="center">(63)</div>

V1 (COLLECTION of information)
<u>acquisition</u>, <u>assemble</u>, <u>collect</u>, <u>gather</u>.

V2 (CATALOGUING)
<u>catalog</u>.

± V3 (ABSTRACTING)
<u>abstract</u> (texts) <u>condense</u>, <u>make abstracts</u>, <u>summarize</u>.

± V4 (SYNTHESIS)
<u>prepare</u> (literature search) <u>reports</u> <u>analytical</u> (and <u>critical</u>) (research) <u>reviews</u> (literature) <u>surveys</u>.

V5 (CLASSIFICATION and INDEXING)
<u>analyze</u>, <u>catalog</u>, <u>categorize</u>, <u>classify</u>, <u>code</u>, <u>correlate</u>, <u>cross-reference</u>, <u>describe</u> (documents), <u>express... in a neat fashion</u>, <u>index</u>, <u>order</u>, <u>organize</u>, <u>reduce...to a standard form</u>, <u>represent</u> (content), <u>produce</u> classification information.

V6 (EXTRACTING)
auto-encode, auto-index (or self-), extract, index,
produce a bibliography (a concordance, a KWIC-index).

± V7 (TRANSLATING)
translate.

± V8 (CODING)
abstract (into punched-cards) (?) code, notation,
translate (into punched-cards) (?)

V9 (STORING)
arrange (in a file), bank, code (on punched-cards) (?),
file, place (in a store or file), store.

V10 (SEARCHING)
explore, interrogate, produce (a search), scan, search,
sort (cp. "diminishing false drops and blank sorts",
J 413).

V11
find, locate, retrieve, select.

V12 (a) (DISSEMINATING)
display, (selective) dissemination, present.

(b) bring (information to scientists)
communicate, convey, display, distribute,
disseminate, document, impart, inform, get
(information) into the hands of, keep ... informed,
present, provide, sell, serve, service.

(c) (+ need for information, etc., as Object.)
fulfil, satisfy, serve.

(d) (+ information as Object.)
make available, accessible, offer, provide.

V13 :(with scientist as Agent or Dative, and Goal.)
borrow, collect, gather.

V14 :(with scientist as Agent or Dative.)
have access to, accessible, available, control, find,
get, keep informed, keep up with, learn of, obtain.

V15 :(with scientist as Agent.)
use, read (document).

In view of what precedes, I would argue that the linguist
can only resort to data collection and to a metalanguage if he
wishes to set up a canonical list of the predicates of fields
such as documentation. In other words, I have the feeling
that, considering the complexities of such a field, any
linguistic theory currently available is unable to predict
the wealth of lexical possibilities that would need to be
predicted. Instructive as they may be, monographs of the
type Fillmore has presented on small sets of verbs ('Hitting'
and 'Breaking', 1967; 'Verbs of Judging', 1969b) do not
provide theoretical tools powerful enough for dealing with a
field like documentation. Besides, even if more sophisticated
theories were available, I believe that semantic primes that
are relevant to pragmatic competence in a field like documenta-
tion would still have to be given by a metalanguage set up
a priori by experts, in the form of stipulative definitions.

3.4. ASYMMETRIC CONJUNCTION AND CASE.

3.4.1. Towards a case representation of asymmetric conjunction.

In this chapter, I wish to present an analysis of
asymmetric conjunctions of sentences and relatable construc-
tions in terms of an extension of case theory. In the present
section, I shall present the type of representation which I
believe to be needed. Before doing this, I shall consider
a fairly simple example of asymmetric conjunction in order
to identify the types of constructions that are relatable
to it. I realize that some of the paraphrases presented below
may sound a bit odd. However , I think that they are logical-
ly possible and that a general methodological principle of
TG theory, which I believe to be most fruitful and revealing,
justifies an examination of all the possible constructions
relatable to asymmetric conjunction. According to this
principle, sentences are derived from their most complete
representations or, to put it differently, the paraphrases
of a sentence that most explicitly reveal the relations
between the sentence constituents are considered to be closer
to deep structure than the less explicit paraphrases.

In 3.3.3., I have pointed out that next to conjunction
('storage AND retrieval of information'), various types of
constructions can serve to express the concept of documenta-
tion or relatable concepts : for instance, expressions of
purpose (store information IN ORDER TO retrieve it),
instrumental expressions (retrieve information BY storing it),
relative clauses and transformationally relatable construc-
tions (retrieval of the storED information) and expressions
of result (the storage of information RESULTS in its retrieval).

In order to carry my discussion of asymmetric conjunction a
step further I shall go out from just one example which is
given below in (1), and which is a simplified version of a
sentence occurring in J 1 :

(1)

The Center will contribute to the development of science
by collecting, processing, storing, and retrieving
scientific information.

The following points can be made about this sentence.

(1) The gerunds in (1) can be conj ned by AND BY :

(2)

The Center will contribute to the development of science
BY collecting AND BY processing AND BY storing AND BY
retrieving scientific information.

(11) (1) can be converted into the conjunction given in (3) :

(3)

The Center will collect, process, store and retrieve
scientific information and contribute to the development
of science.

(111) (1) can be converted into a sentence containing an

expression of purpose :

(4)

The Center will collect, process, store and retrieve
scientific information IN ORDER TO contribute to the
development of science.

(iv) (1) can be converted into a sentence containing both

an 'instrumental' expression and an expression of purpose,

as in (5) (a) which is modelled after the attested

sentence quoted in (5) (b) :

(5)

(a) The Center will carry out its PURPOSE to contribute
to the development of science BY collecting, process-
ing, storing, and retrieving scientific information.
(b) Aslib carries out its purpose to develop the practice
of special librarianship (...) by means of programmes
of education and research, (1) by educating...,(2)
helping... (J 1102).

Furthermore, (1) can be converted into a sentence in which all the verbs depend on "purpose" :

(6)

The Center will carry out its PURPOSE to collect, process, store, and retrieve scientific information and to contribute to the development of science.

To me, however, a sentence such as (6) seems to be truncated version of the type of sentence illustrated above in (5) in that some 'instrumental expression (by ...) seems to be implicit in it.

(v) A purpose expression can be ... inside an 'instrumental' expression (7) (a), and an 'instrumental' expression can occur inside a purpose expression (7) (b) :

(7)

 (a) The Center will contribute to the development of science by collecting information in order to process, to store, and to retrieve it.
 (b) The Center will collect information in order to contribute to the development of science by processing, storing, and retrieving the information.

(vi) (1) can be converted into a sentence with "result" as its main verb :

(8)

The Center's collecting, processing, storing and retrieving of scientific information will RESULT in the Center's contributing to the development of science.

(vii) Relative clauses or transformationally related constructions can be introduced into (1) :

(9)

The Center will contribute to the development of science by collecting information, by processing the collected information, by storing the processed information, and by retrieving the stored information.

(viii) Conjunction shows up in all the sentences just given (1 to 9) in the form of and or of a comma, or of both.

In all the linguistic literature with which I am
acquainted, I have found no rule or system of rules that would
be able to generate semantic descriptions for the sentences
just given. Typically, Ross (1967, 169) does not attempt to
characterize the purposive relationship which he believes
should hold between conjoined sentences such as (10) (a)
for them to be convertible into a noun-phrase with relative
clause such as (10) (b) :

(10)
(a) I went to the store and bought some whisky.
(b) The whisky I went to the store and bought.

Now, what strikes me about the sentences given in this
section is this :

(1) It seems that they fall basically into two types of
surface constructions. In the first type, we have just
conjunction of VP's : The Center VP1 and VP2 and VP3 etc.
(as in 3). The second type consists of a sentence in which
we have both conjunction and some other type of relation
(e.g. purpose), and the two types of construction seem to
be complementary : for instance, The Center VP1 IN ORDER TO
VP2 AND TO VP3.

(ii) The sentences of the second type can in turn be subdivided
into two types, according to the nature of their main
verbs. In the first subtype, the main verb can be described
semantically as an abstract predicate : for instance,
The purpose of— .. .. results in—. In the
second subtype, the main verb . one of those verbs
described in 3.3.4. as v.. etc. : for instance,
'the Center V1s... in order to V2... and ...' or 'the Center
V2s... by V1ing...'.

(iii) In both sentence types men.... ir ...). there always
seems to be a division into two parts. the clearest
examples are structures of the type 'x Vs this (±and...)
IN ORDER TO V this (± and...)' or of the type 'x Vs this
(±and.. BY Ving this (±and...)' , where IN ORDER TO and BY
mark the partition.

If the above observations are correct, the fundamental problem lies, I think, in adapting the well-known rule schema $S \longrightarrow C\ S\ x$ (sentence rewritten as a conjunction of any number of sentences) in such a way that the relatedness between conjunctions of sentences and other, partly equivalent, sentence structures can be captured.

The solution I propose is the following. It consists of resorting to a form of representation in which a sentence (S0) is rewritten as a pair of case labels which themselves are rewritten as a pair of sentences (S1 and S2). The system is recursive in that it allows for S1 or S2 or both to be rewritten as a pair of sentences (S1', S2') and for these sentences themselves to be rewritten (as S1'', S2''), etc. Following a recent proposal by Fillmore (1970b) S1 and S2 are rewritten directly from case labels without any intermediary rewritings (such as the case marker K + NP in Fillmore 1968; moreover, as stated in 3.1.1., I shall write S for what Stockwell et al. represent as NOM and rewrite as N + Cases : Stockwell et al. 1968, PS rule 7, page 34). The case labels which I propose are Cause for S1, and Goal for S2, and are intended to represent the fact that the event or state of affairs described in S1 is regarded as a necessary logical condition for the event described in S2 to be possible. Cause is a generalization of a "role" label proposed by Langendoen. He argues that case (or "role") theory needs to state a distinction between two situations. When an "instrument-like expression" is used in a sentence with an Agent expressed or understood, Langendoen proposes to analyze the expression as Instrument. When, on the other hand, an instrument-like expression is "used in a sentence without an understood Agent", he proposes to analyze it as Cause : for instance, "loss of blood" in "loss of blood killed the victim". Langendoen, I think correctly, points

out that a sentence "Poison killed the victim" is ambiguous
with respect to the absence or presence of an understood
Agent, and he proposes to capture this by having a role or
case structure consisting of either Instrument and Agent
or just Cause (Langendoen 1970, 74-5). I borrow the Goal
label from Fillmore who has recently proposed to generalize
the definition of the Goal case for it to include his former
Result case. Moreover, I propose that either S0 alone, or
S1, or S2, or both, can be marked + main, to represent the
fact that any of the sentences in question can be selected
as the main sentence (in a paper which has not been available
to me, Postal makes use of a "-main" convention to derive
restrictive relative clauses : Postal 1967, in Annamalai 1970,
140).

In addition to the + main-convention, I propose a three-
valued system of features which are attached to the case
labels : for instance, the features [+ Reason] and [+ Means]
can be attached to the Cause label, and features such as
[+ Purpose] and [-Result] to the Goal label. I suggest that
the value ±, standing for 'irrelevant', is applicable to
conjoined constructions, since such concepts as result and
purpose are left unexpressed in conjunctions. To sum up, the
notation system proposed is intended to symbolize both the
similarities and differences between asymmetrically conjoined
sentences and the various relatable constructions as illustrated
at the beginning of this section · discussed in 3.3.3.

I shall now illu. . . . · em I propose by presenting
a few types of situat.... that arc . .dictable on the basis
of such a system. Analyse of e r.. .. attested examples
will be presented later, ... ns ⁿ · ... and 3.4.4. In
these analyses it will be necessar, to a.. few features
to those mentioned so far. In the examples I present below,

I assume the correctness of the "conjunction reduction" hypothesis (Stockwell et al. 1968, 323). What will be said in the next section on projection and relative clauses (3.4.2.) can serve as an argument for this assumption.

(1) In the following sentence (11), SO is marked + main because the sentence has an abstract predicate ('result in') as its main verb. The Cause-clause (S1) is the subject in the surface form of the sentence, as is "Loss of blood" in Langendoen's sentence "Loss of blood killed the victim". I have attached the negative feature specifications [-Means] and [-Reason] to S1 and [-Purpose] to S2, because it seems to me that no such concepts are involved. Even though it is redundant with the main verb, a feature [+ Result] is attached to S2, because such a feature is needed elsewhere. S1' and S2' are marked + main to represent the fact that, since they are conjoined, they are on the same footing.

(11)
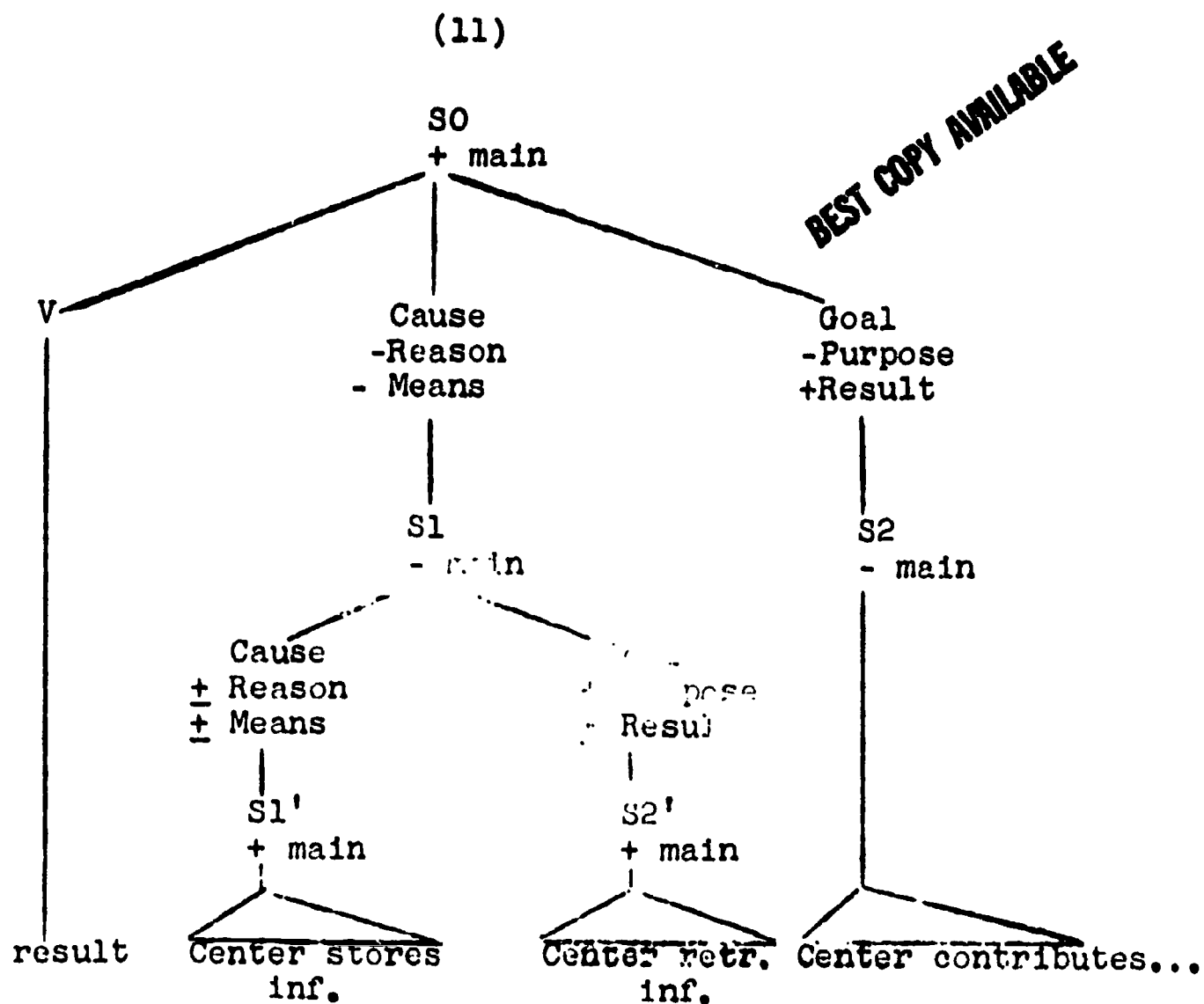
> Example : The Center's storage and retrieval of
> information results in the Center's
> contributing to the development of
> science.
>
> Cp. : The Doctor's speedy arrival and careful
> examination of the patient resulted in
> her speedy recovery.

(11) In sentence conjunction, I propose to mark both S1 and
S2 as ± main. In (12), all the other feature specifications
are marked ± : this is how I propose to represent the
fact pointed out   earlier that conjunction blurs or
obscures relations that are expressed in other, parallel,
constructions. I suggest, however, that this need not
always be the case in sentence conjunction or coordination.
Thus, S2 would be marked [+ Result] to represent the
following sentence (cp. the second sentence given below
in 12) : 'The doctor arrived rapidly and examined the
patient very carefully; (and) as a result, she soon
recovered'.

(12)

SO
- main

Cause
± Reason
± Means

Goal
± Purpose
± Result

S1
+ main

S2
+ main

Cause
± Reason
± Means

Goal
± Purpose
± Result

S1'
+ main

S2'
+ main

Center stores inf.   Center retr.inf.   Center contributes to...
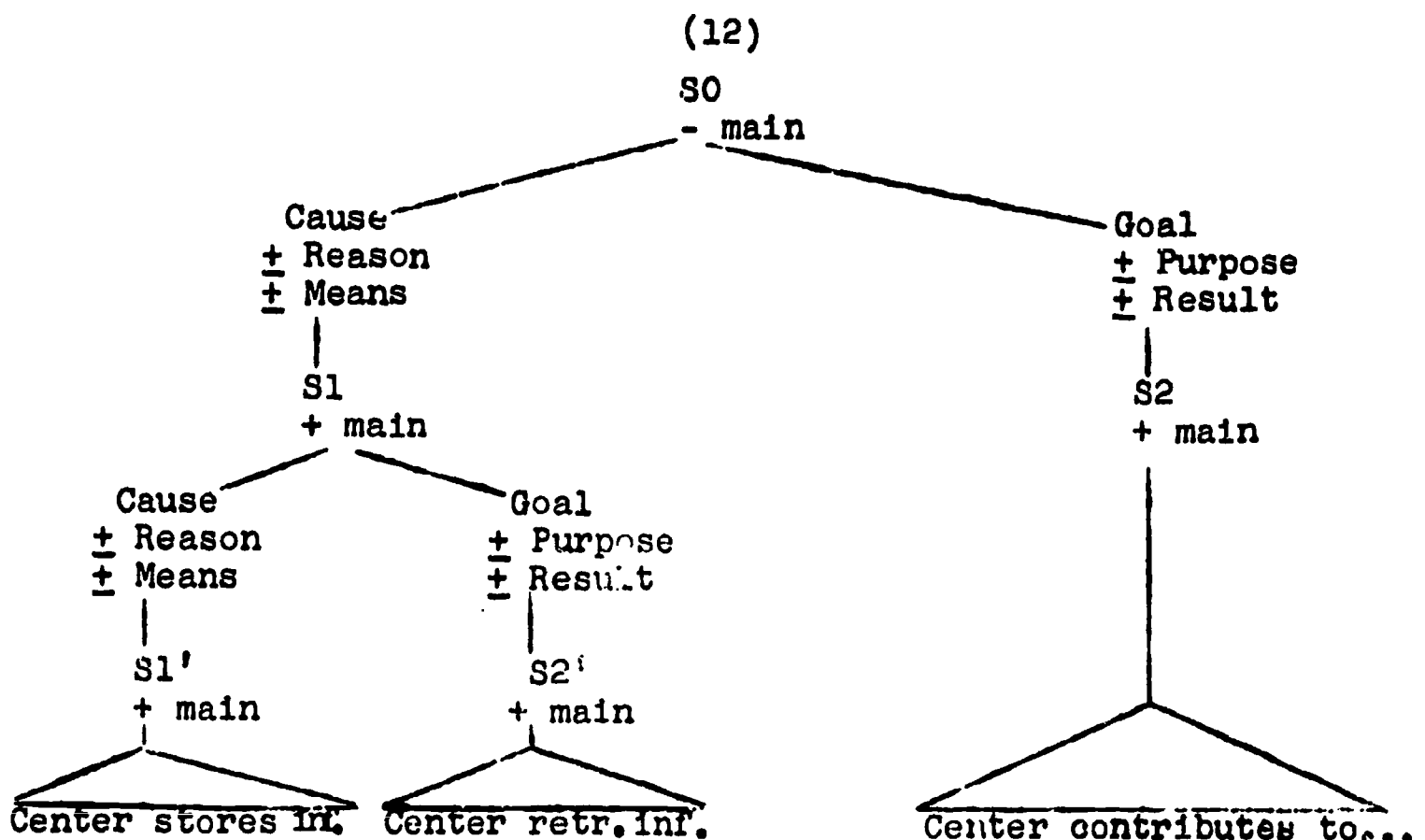
Example : The Center stores and retrieves information
and contributes to the development of
science.

Cp. : The Doctor arrived rapidly and examined the
patient very carefully, and she soon recovered.

(111) The representation which I propose for by-expressions
which Fillmore 1970b would :· ·:··:esent as Instruments
is given below in (13). The sentence with the main verb
is marked ± main and the by-expression is assigned
the feature [+ Means], for lack of any better name.
A feature [+ Reason] rather than [+ Means] seems
appropriate to represent the because-clause in the other
example given below (cp. Hornby 1954, 240 f.).

(13)

SO
- main

Cause
-Reason
·-Means··

Goal
-Purpose
-Result

S1
- main

S2
+ main

Cause
± Reason
± Means

Goal
± Purpose
± Result

S1'
+ main

S2'
+ main

Center stores inf.   Center retrieves inf.   Center contributes to...
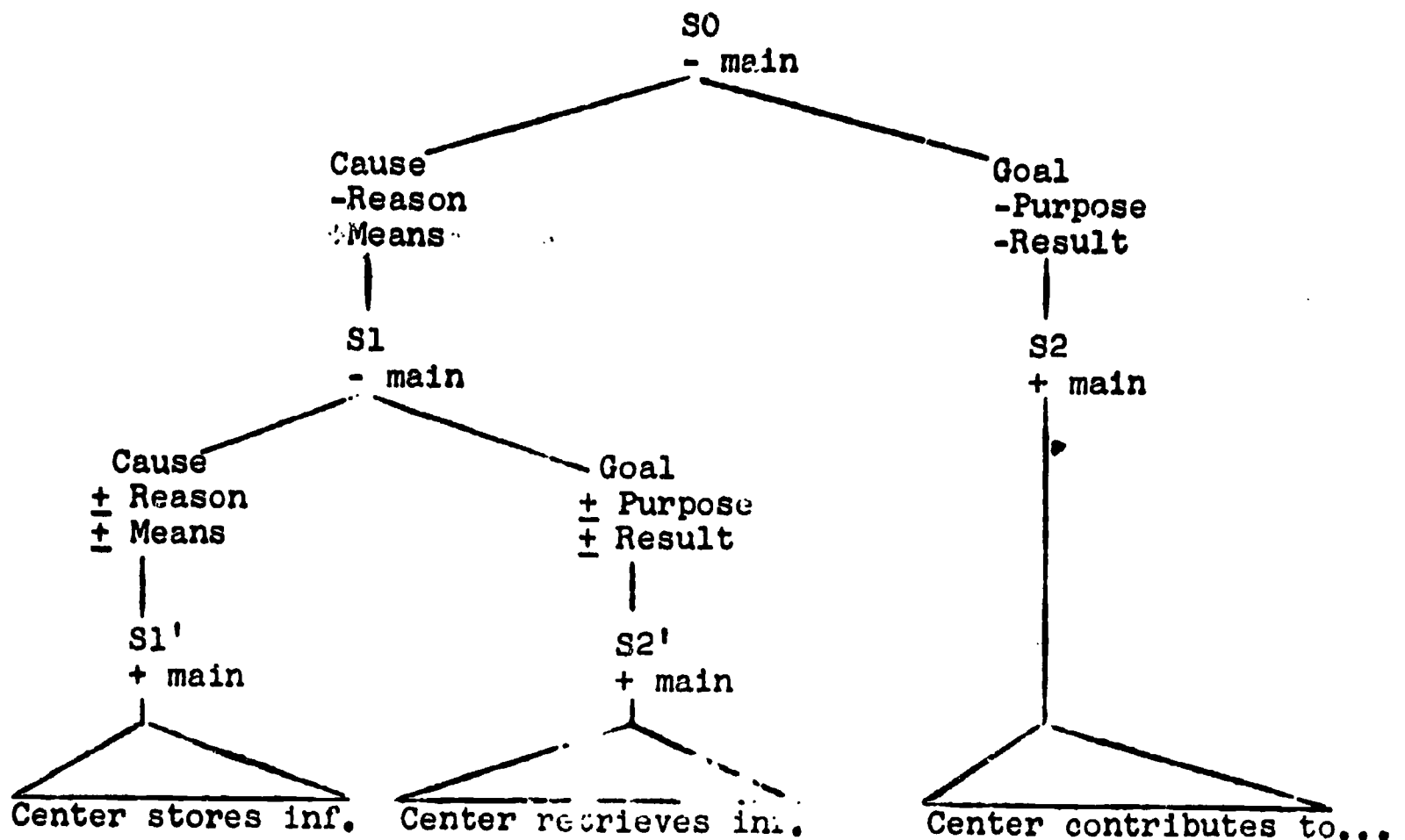
Example : The Center contributes to the development
of science by storing and (by)retrieving
information.
Cp. : The patient soon recovered because the Doctor
arrived rapidly and examined her very carefully.

(iv) Purpose expressions corresponding to (iii) would be
analyzed as in (14). The expression of purpose is marked
- main and [+ Purpose]

(14)
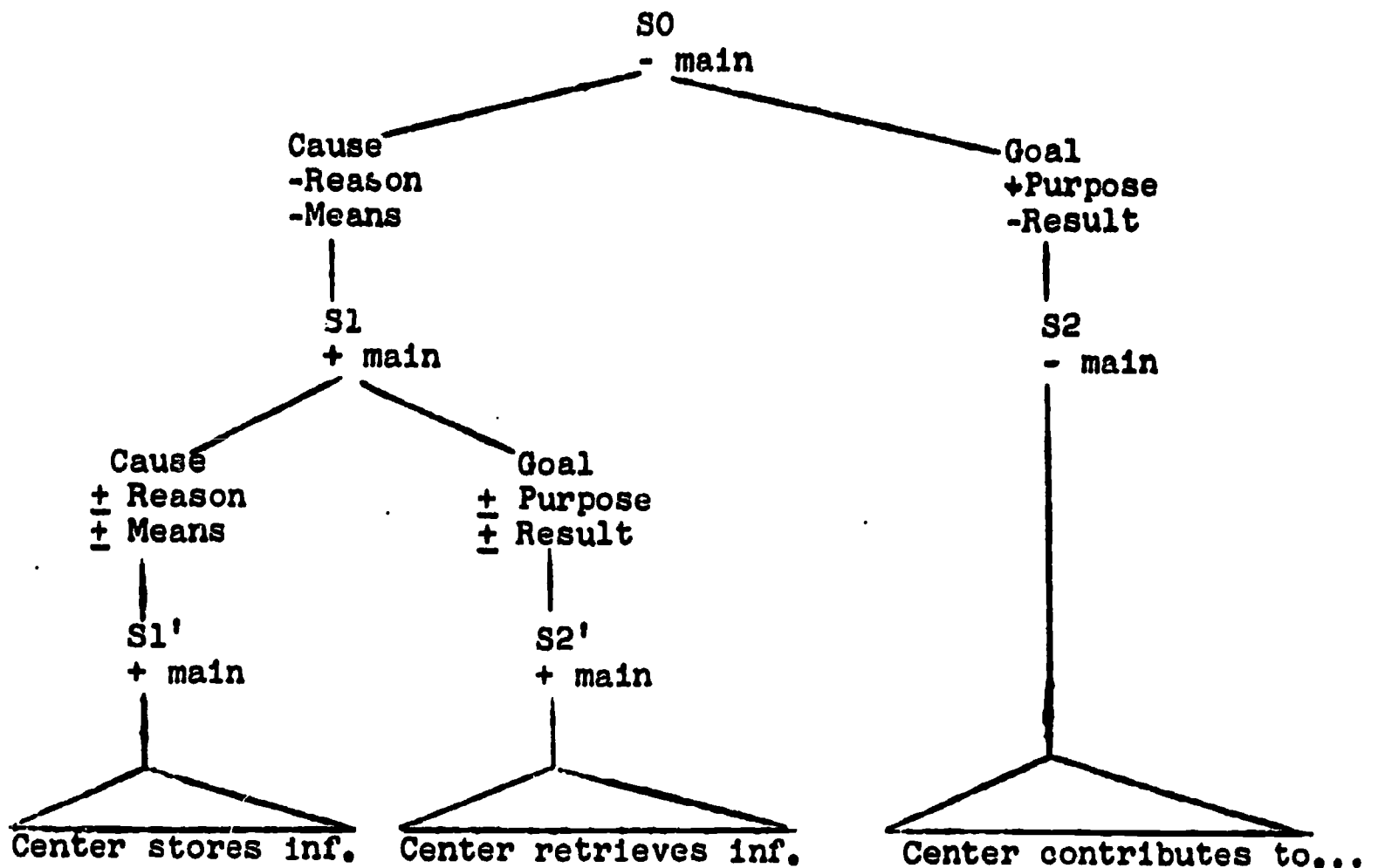
Example : The Center store. .nd retrieves information
in order to contribute to the development
of :. .oc
Cp. : The Doc:..r arr.vuu very quickly and examined
the patient very carefully in order that she
might soon recover.

The main advantage which I would claim for the type
of representation I propose has already been mentioned :
it captures at the same time the similarities and differences
between asymmetrically conjoined sentences and the various
constructions relatable to it. Note, in particular, that the
recursiveness of the system makes it possible to represent
'mixed' constructions : for instance, the occurrence of
a purpose phrase in a by-phrase.

Another advantage is that the concept of Vlast proposed
in 3.3.4. can now be more precisely defined. Thus, I suggest
that what I have called Vlast can now be defined as the
main verb of any sentence   which is an exponent of the
Goal case.

Note that, under the proposals just made, a given
lexical item could have more than one description. For
instance, the verb retrieve would have different analyses in
(12) above and in the following diagram (15). I regard this
consequence as by no means unacceptable, but on the contrary
as a desirable one :

(15)

```
                              S0
                            - main

        Cause                              Goal
        ± Reason                           ± Purpose
        ± Means                            ± Result

          |                                  |
          S1                                 S2
          + main                             + main

                              Cause                    Goal
                              ± Reason                 ± Purpose
                              ± Means                  ± Result

                                |                        |
                                S1'                      S2'
                                + main                   + main


    Center stores inf.      Center retrieves inf.   Center contributes to...
```
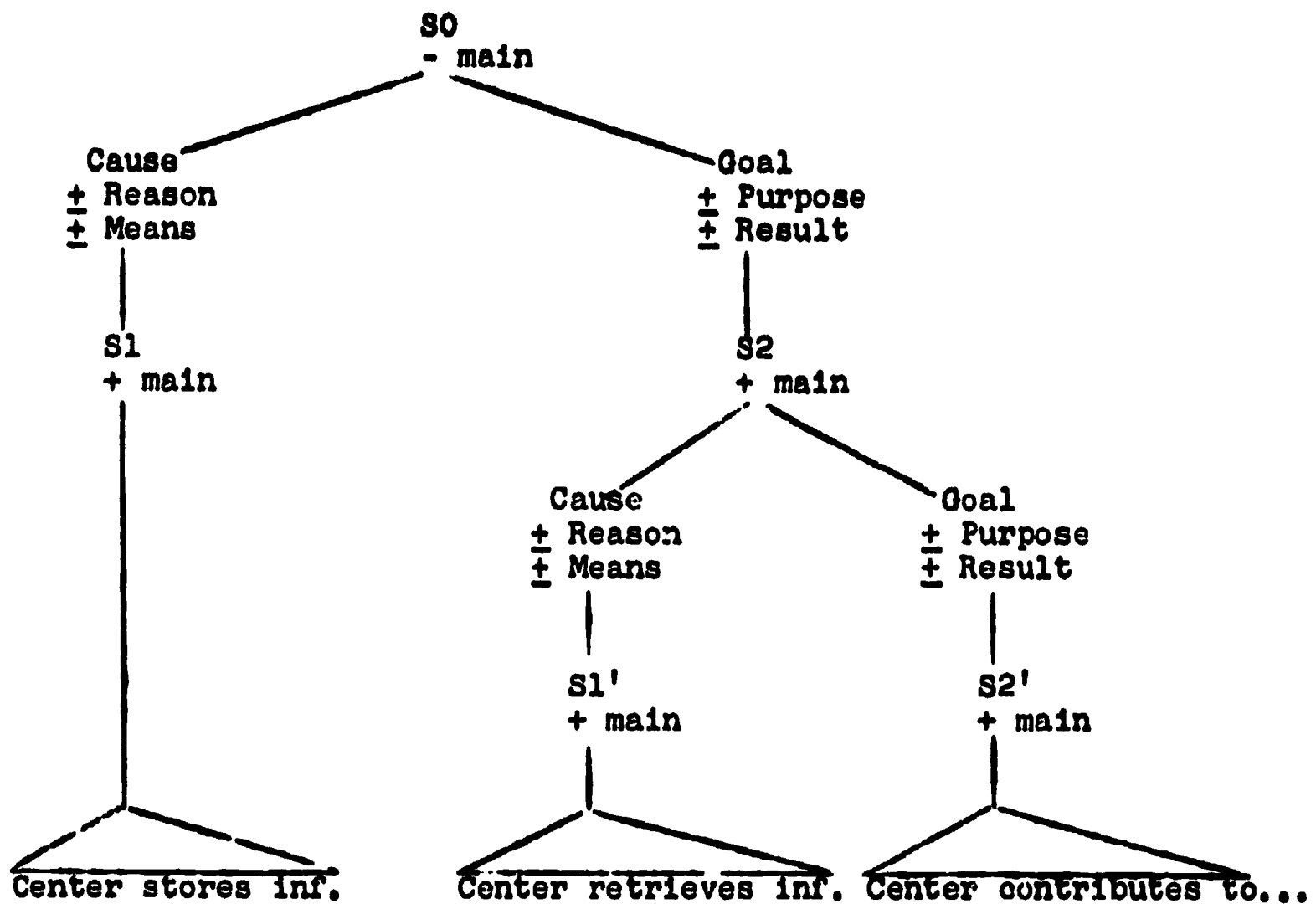
Example : The Center stores information, and it
          retrieves this information and contributes
          to the development of science.
Cp. : The Doctor arrived rapidly; he examined the
      patient carefully and she soon recovered.

3.4.2. Langendoen's projection and relative clauses.

The only important type of construction relatable to
asymmetric conjunction that was left undiscussed in the
preceding section is relative clauses, and the various
constructions derivable from them : for instance, 'retrieval
of the information which has been processed and stored' or
'retrieval of the information stored and processed' as
related to 'process, store and retrieve information'. In
what follows, I shall not be concerned with questions of
tense : e.g. 'has been' in the above examples. What I wish
to do now is to argue for generaling relative clauses on
the basis of what Langendoen 1967 calls projection, namely
the process by which an argument receives specifications
corresponding to the predicates with which it has contracted
a relation in previous discourse. Moreover, on the basis
of the analyses presented in 3.4.1., I shall argue that
projection needs to operate irrespective of the type of
construction in the previous discourse, that is, no matter
whether it is asymmetric conjunction or some other, relatable,
construction (cp. J 985 : "procedures used to store the
punched card information... this stored information...").
The whole of the present discussion is based on the method-
ological principle already mentioned that sentences are
to be derived from their most complete representations.

(1) If the semantic representations proposed in 3.4.1. are
correct, sentence conjunction can no longer be regarded
as the ONLY source of relative clauses. In other words,
relative clause formation would be just a matter of
projection and would not be sensitive to surface
structure (such as conjunction), but only to deep
structures (such as those just proposed). The frequent
occurrence of deictics in NP's relatable to relative
clauses (see 16) is a strong argument for the central
place I propose for projection in relative clause forma-
tion.

BEST COPY AVAILABLE

Compare (16) (a), (b), (c), and (d) :

(16)

(a) The Center _stores_ information AND retrieves
the _stored_ information.
(b) The _storage_ of the information RESULTS in the
retrieval of _this_ stored information.
(c) The Center _stores_ information IN ORDER TO
retrieve this _stored_ information.
(d) BY _storing_ information, the Center (is able
to) retrieve _this stored_ information.

(11) (16) (b) is a case in which the deletion of the source
sentence is most obviously RULED OUT : x ∅ results in
the retrieval of the stored information.

(iii) An example such as (16) (d) suggests that relative
clause formation must be sensitive to the left-to-right
order of constituents in surface structure. Devices such
as global derivational constraints therefore seem to be
necessary for generating relative clauses. This type of
rule, the formal mechanism of which has been described
by Lakoff 1969, is the only one known to me that is able
to specify, for instance, that a relative clause is only
possible if the 'source sentence' precedes the NP into
which the relative clause can be embedded in the resulting
surface structure. Compare (17) (a) and (b) :

(17)

(a) By _storing_ information, the Center is able to
retrieve this _stored_ information.
(b) x The Center is able to retrieve this _stored_
information by _storing_ it.

(iv) I now wish to turn to some more abstract surface expressions than those so far discussed. Such expressions obviously need to be related to relative clauses.
I believe that a distinction needs to be made between two superficially similar constructions. One type is exemplified below in (18), the other in (19).

(18)

   (a) ... a regular scrutiny of over one thousand current periodicals with consequent indexing and distribution of the <u>resulting</u> information (J 466).

   (b) ... how to conduct bibliographical investigation and how to present the <u>resulting</u> data (J 142).

(19)

   (a) the increasing volume of literature, the rapid fragmentation of biology and <u>its resultant</u> increase in semantic problems (J 931).

   (b) Although subjects exhibited some reliability in selecting representative sentences, the <u>resultant</u> reliability was low (J 816).

I believe the difference between the examples of (18) and (19) to be the following. "The <u>resulting</u> information" in (18) (b) is equivalent to 'the information that resulted from the events or actions previously mentioned'; for instance, "indexing and distribution of the <u>resulting</u> information" is equivalent to 'indexing and distribution of the <u>indexed</u> information'. I regard this as an instance of projection. I suggest that the other type is to be analyzed along the lines proposed in the preceding section : compare, for instance,'The Doctor's speedy arrival and careful examination of the patient and the <u>resultant</u> improvement in her condition'. This 'resultant' could be generated optionally, given the presence of a feature [+ Result] attached to the Goal clause.

Also note that in any case it cannot be 'derived' from a previous conjunct in the fashion proposed by Annear and Langendoen.

(v) In this discussion of relative clause formation and projection, I have to mention a further complication. I have found no attested example of the constructions I wish to present in this paragraph, but I believe they are possible. What they suggest is that the various constructions relatable to asymmetric conjunction that were encountered in main clauses are also possible in relative clauses. Consider the following sentences :

(20)

  (a) The Center will contribute to the developmer͟ of science by collecting information, by processing the collected information, by storing the information which has been collected AND processed, and by retrieving the information which has been collected, processed, AND stored.

  (b) The Center will contribute to the development of science by collecting information, by processing the collected information, by storing the information which has been collected FOR processing, and by retrieving the information which has been collected and processed FOR storing.

  (c) ? The Center will contribute to the development of science by collecting information, by processing the collected information, by storing the information which can be processed BY being collected; and by retrieving this information which can be stored BY being collected and processed.

  (d) The Center will contribute to the development of science by collecting information, by processing the collected information, by storing the information whose collecting RESULTS in its being processed, and by retrieving this information whose collecting and processing RESULTS in its being stored.

I regard the possibility of there being sentences such
as those just given as further evidence upholding the
semantic representations proposed in 3.4.1. as well
as the suggestions I have advanced for deriving
relative clauses on the basis of Langendoen's projection
rather than according to the Annear-Langendoen hypothesis.

(vi) Conjunction reduction should operate after the generation
of deictic noun-phrases, and these, in turn, should be
generated after the operation of Langendoen's projection.
On the basis of what has been said in 3.3.2. we can
assume that the order of generation illustrated below in
(21) is the one that needs to be adopted (I ignore the
problem of the first noun-phrase "mechanized information
retrieval systems") :

(21)

(a) Most mechanized information retrieval systems depend
upon extracting information from documents,
processing the <u>extracted</u> information in some manner
and finally placing the <u>extracted and processed</u>
information in a file so that the information
<u>which has been extracted, processed, and filed</u>
can be retrieved by machine.

(b) Most mechanized information retrieval systems
depend upon extracting information from documents,
processing <u>such</u> information in some manner and
finally placing <u>this</u> information in a file so
that <u>the</u> information can be retrieved by machine.
(J 321).

(c) Most mechanized information retrieval systems
depend upon extracting, processing and filing
<u>information</u> so that <u>it</u> can be retrieved by machine.

(vii) Further evidence for regarding projection rules that
create relative clauses as very early rules is that
projection may be needed to explain the occurrence of
certain lexical items, as has been argued in 3.3.2.
Consider, for instance the expression <u>"on deposit"</u> in :
"<u>a full account of his research</u> (...) which <u>is permanently</u>
<u>filed in a central repository</u> (...) the full papers <u>on</u>
<u>deposit</u>" (J 783).

While I am unable to write rules that would generate relative clauses on the basis of projeotion, I would suggest, to conolude, that the possibility which projeotion affords for earlier discourse to be 'mirrored in relative clauses of later discourse is one of the ways in which asymmetrically conjoined sentenoes can satisfy the identity requirements which Z. Harris and R. Lakoff posit for conjoined struotures.

3.4.3. A few attested examples.

The examples analyzed in this section consist of
definitions of documentation, or equivalent expressions.
By equivalent expression, I mean any material which is
equivalent to the word 'documentation' in its global sense:
for instance, 'a center which collects,processes, stores,
and retrieves information' is equivalent to 'a documentation
center'. The analyses will be presented graphically in the
form of tree diagrams such as those presented in 3.4.1.
Under each tree, I give the full quotation of the passage
analyzed. When some material is left unanalyzed in a tree,
it is given between brackets in the full quotation. This
means either that I am unable to analyze the material in
question or that it falls outside the scope of the present
discussion. In the trees given below, C stands for Cause,
and G for Goal, and + for + main and - for - main. Some of
the analyses presented below are no doubt questionable,
irrespective of the general validity of my proposals; the
trouble, however, is that there seem to be no clues either
to refute an analysis or to establish an alternative analysis.
This is most obviously the case with some conjunctions :
see in particular (30) - (32). The only criterion I have
found is the existence of abbreviated expressions, which I
propose to analyze as Goal : "(storage and) retrieval",
"(...and) dissemination", etc. Various words such as
prepositions which can serve as evidence for my analyses are
capitalized in the quotations. Except for -(main), I shall
omit all - and + feature specifications, in order to simplify
the presentation.

(22)
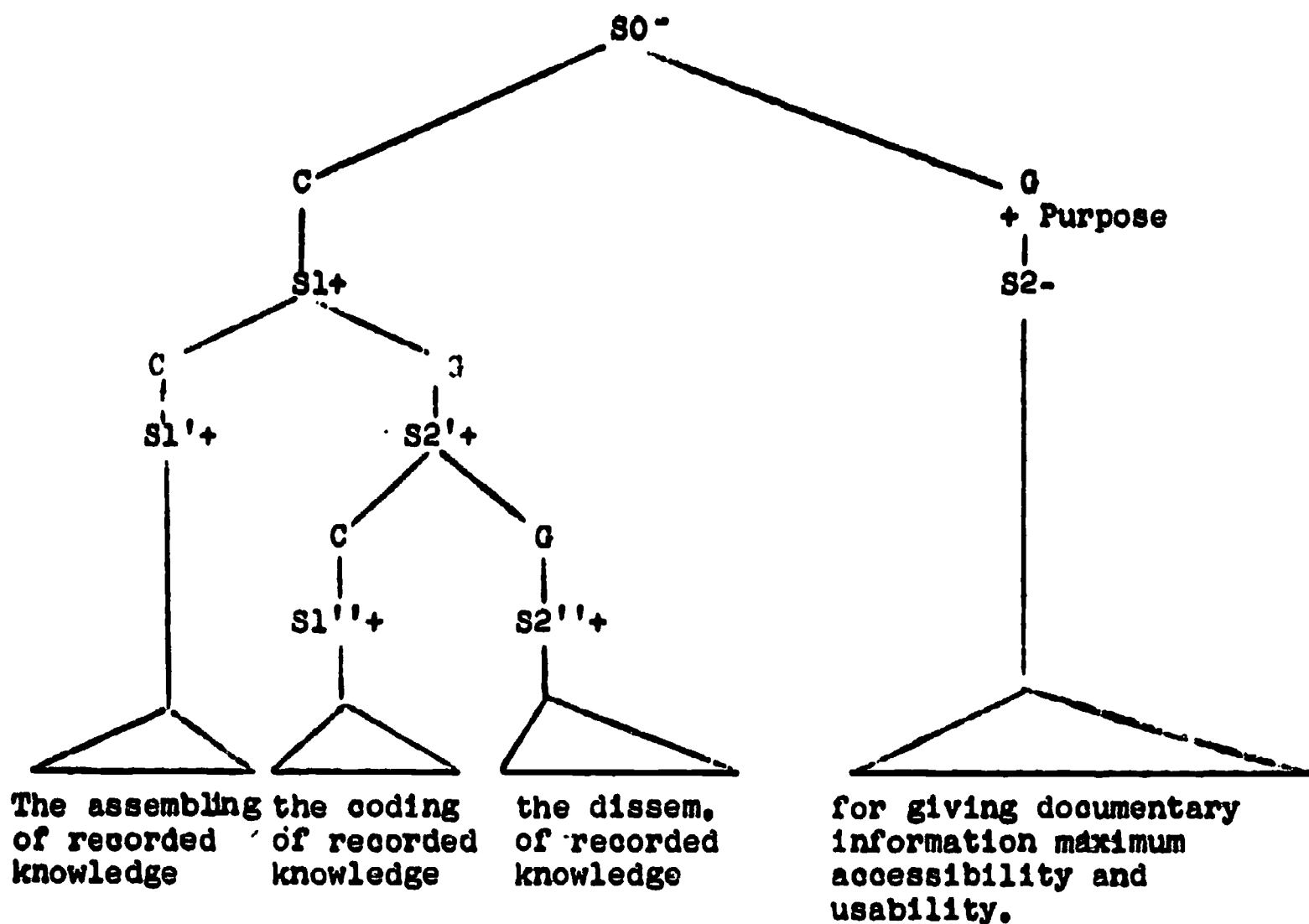
Example : The assembling, coding and dissemination
of recorded knowledge (comprehensively
treated as an integral procedure utilizing
semantics, psychological and mechanical
aids and techniques of reproduction
including microcopy) FOR giving documentary
information maximum accessibility and
usability. (Webster's definition of
documentation).

(23)
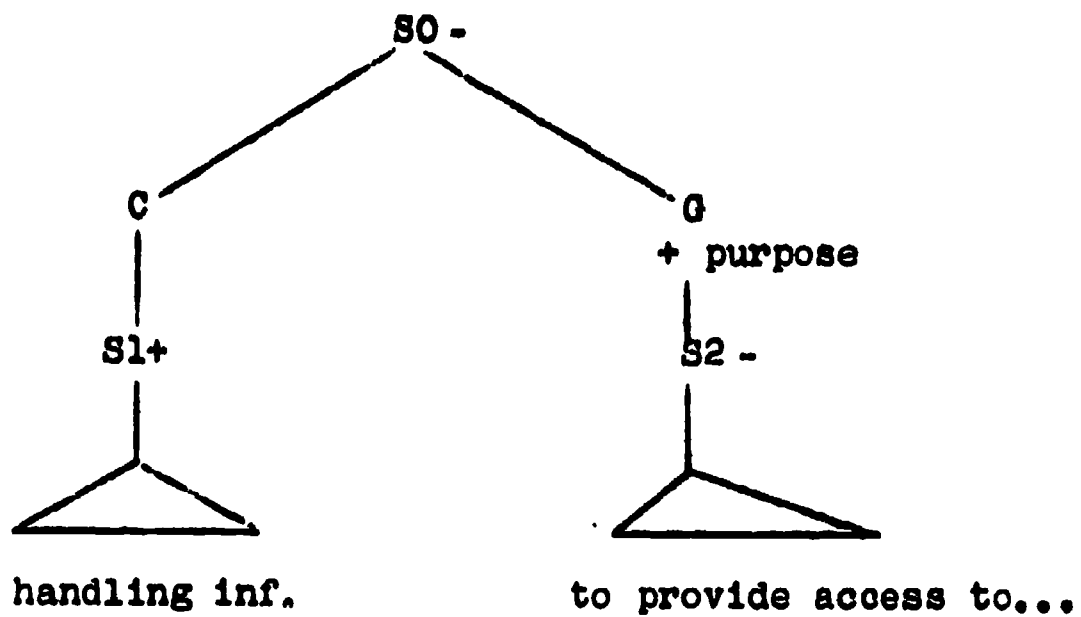
Example : ... handling scientific information
TO provide access to the scientific
literature (J 263).

(24)

SO -

C   G
              + Purpose

S1+   S2-

C   G
    + Time sequence

S1'+   S2'+

C   G

S1''+   S2''+

| extracting inf. from documents | processing such inf. in some manner | and finally placing this inf. in a file | so that the information can be retrieved by machine. |
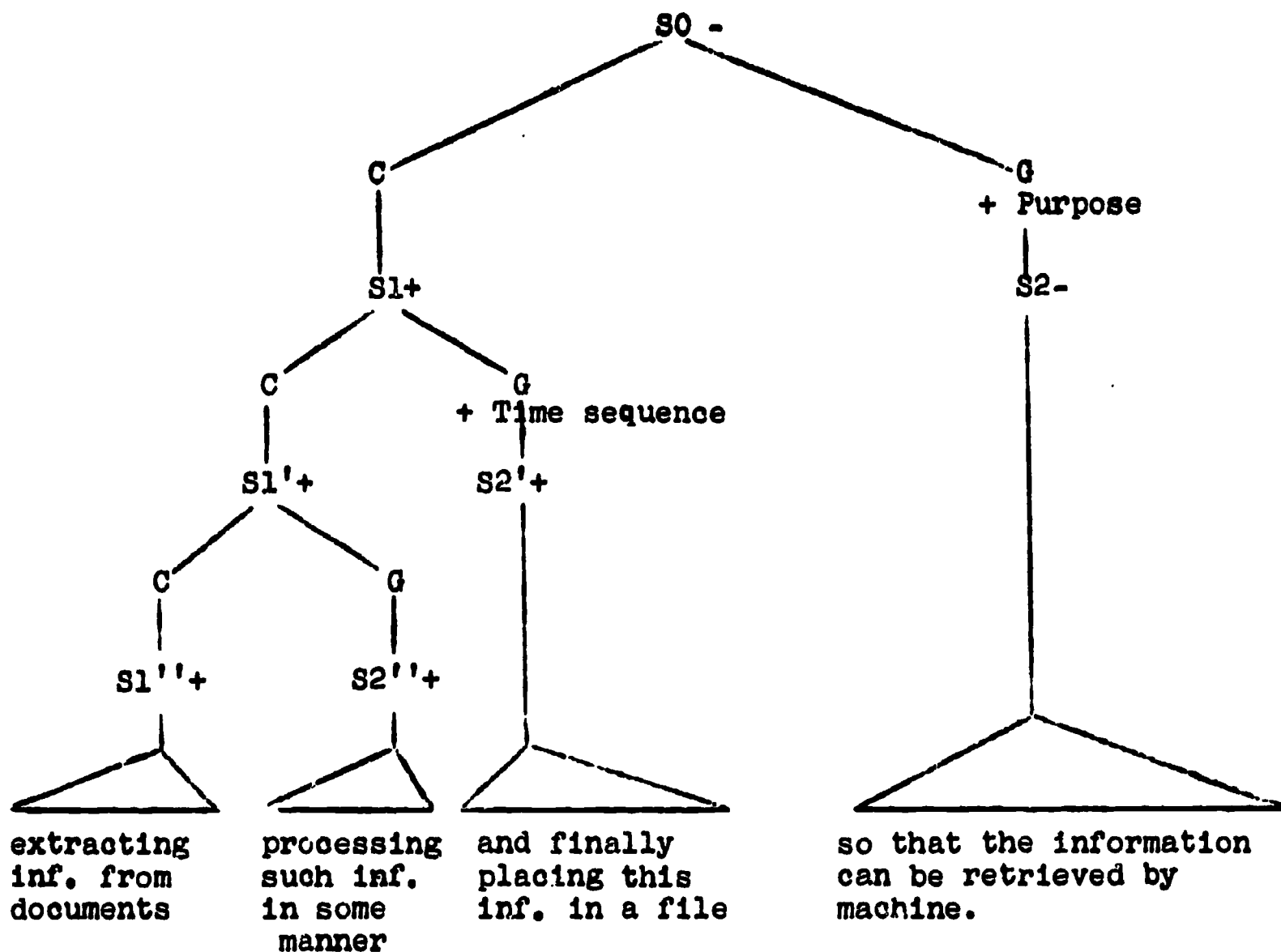
Example :   (Most mechanized information retrieval systems depend upon) extracting information from documents, processing such information in some manner, AND FINALLY placing this information in a file SO THAT the information can be retrieved by machine (J 321).

(25)
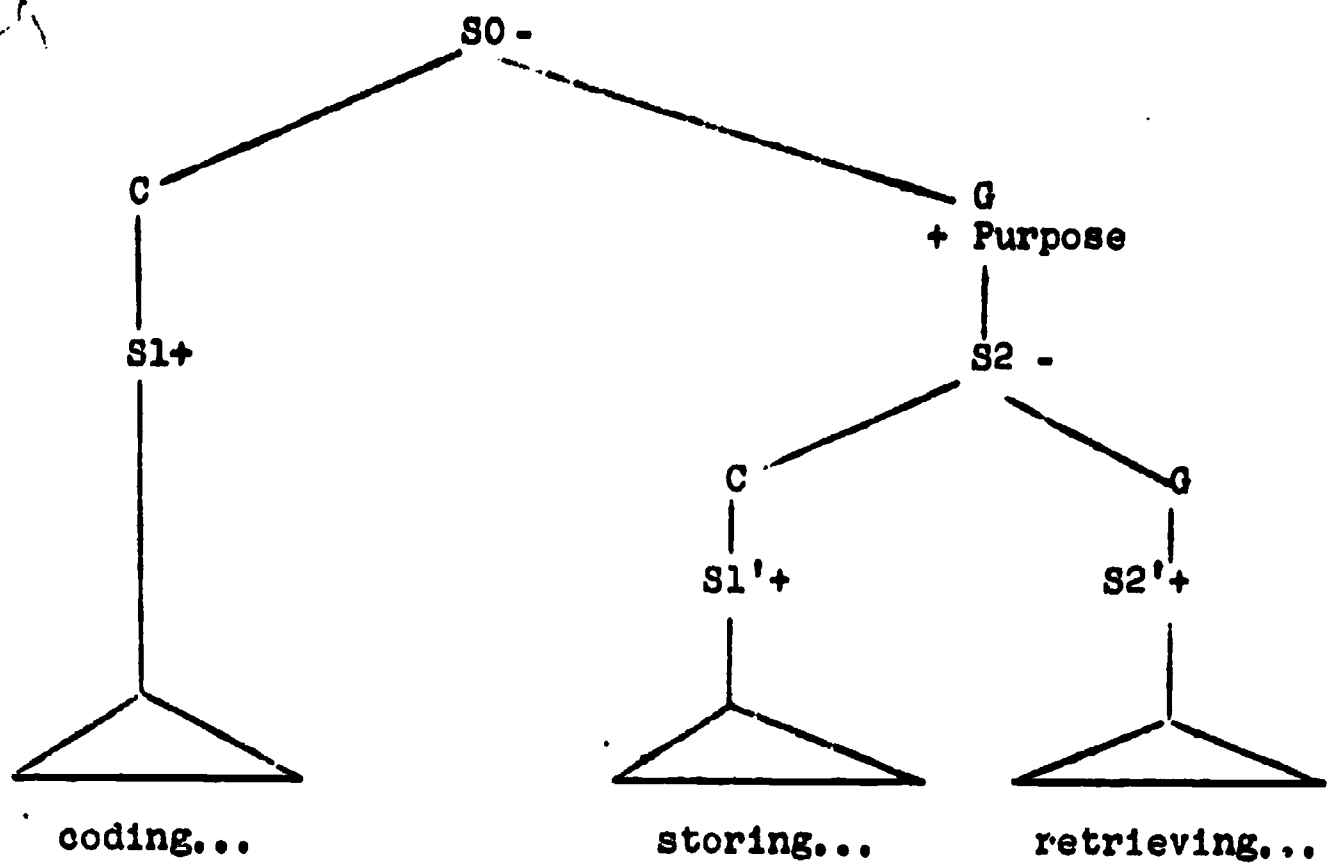
Example : (Introduces a method of) coding two and three
dimensional representations (...) FOR THE
PURPOSE OF storing and retrieving them.(J330)

(26)


```
                              SO.
                    C                       G
                                          + Purpose
              S1+
          C           G                     S2-
        S1'         S2'+

         ∅        institutions          in order to achieve
              ...dissemin...            a net improvement in
              recorded information       scientific productivity
```
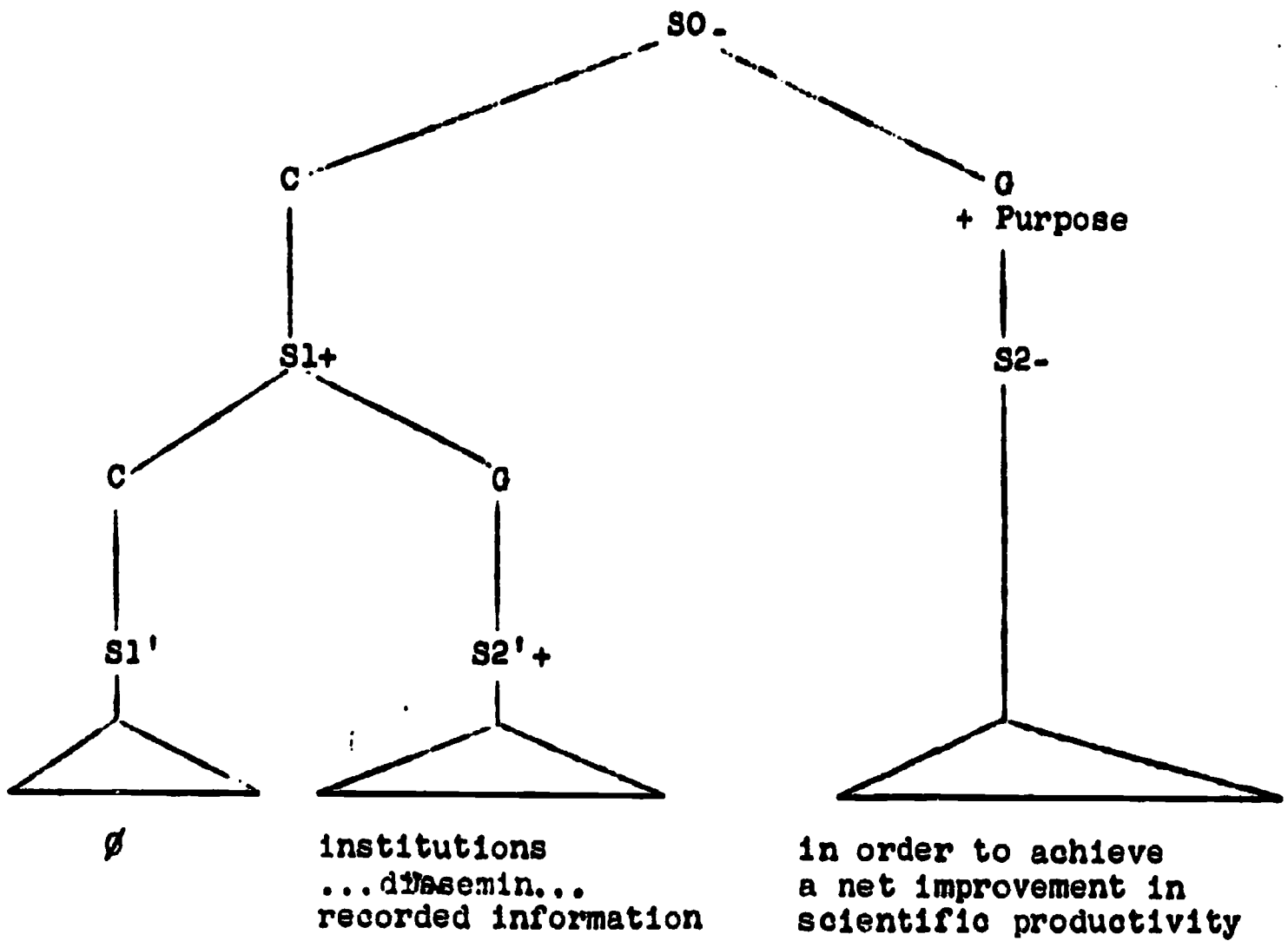
Example :   (institutions that are responsible for)
            the dissemination of recorded information
            IN ORDER TO achieve a net improvement
            in scientific productivity (J2).

(27)

SO -

C

S1+

C

S1'

Ø

increase
the mental
contact
between...

G
+ Purpose

S2-

C

S1'+

so the reader
can proceed
unerringly
and swiftly

O
S2'+

G
+Purpose

S2'-

C

S1''+

to identify
the message
for which..

O

S2''+

to receive
the mes-
sage for
which...
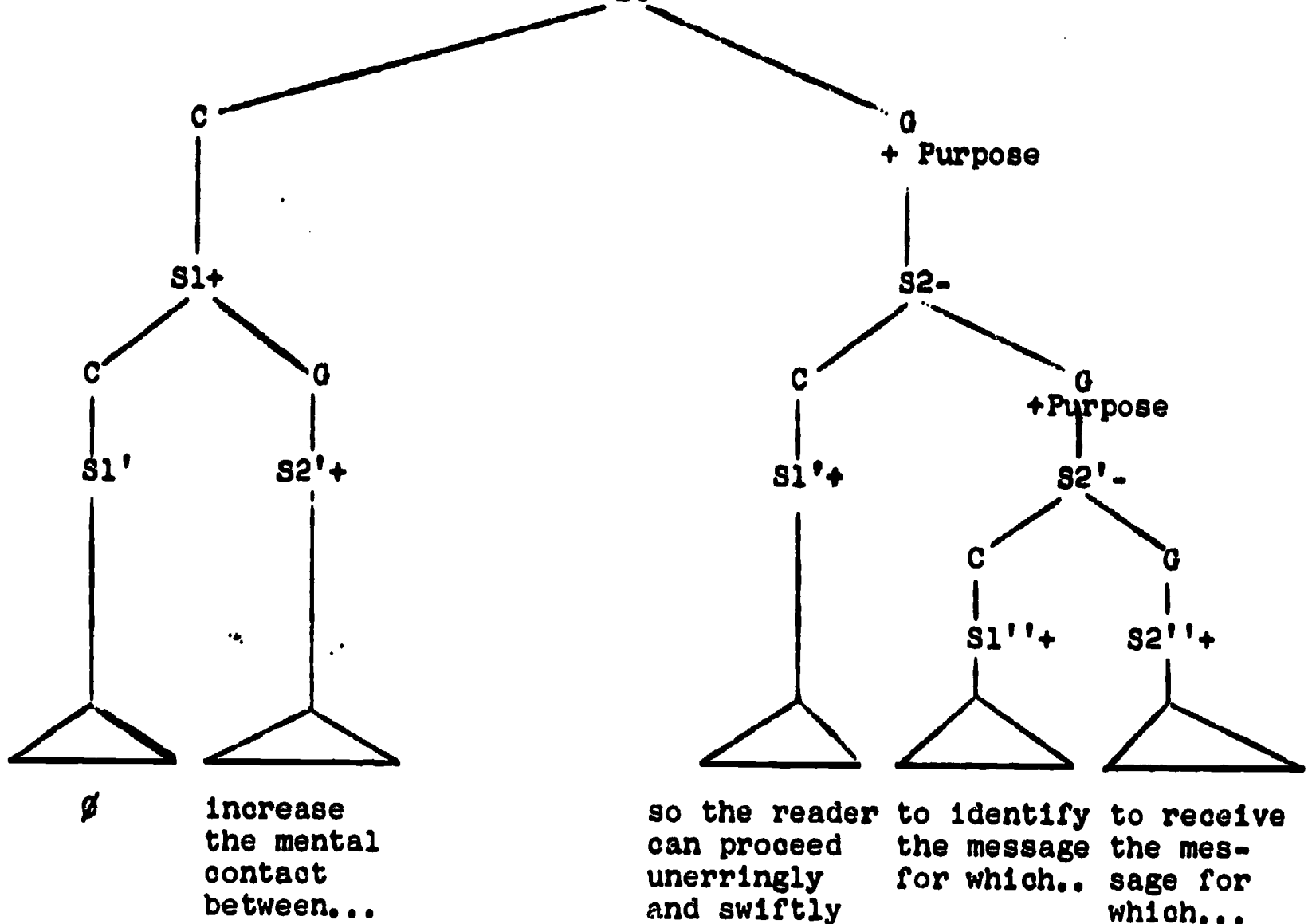
Example : (The basic problem is to) increase the
mental contact between the reader and
the information stored SO the reader
can proceed unerringly and swiftly TO
identify and receive the message for which
he is looking (J257).

(28)

$\varnothing$
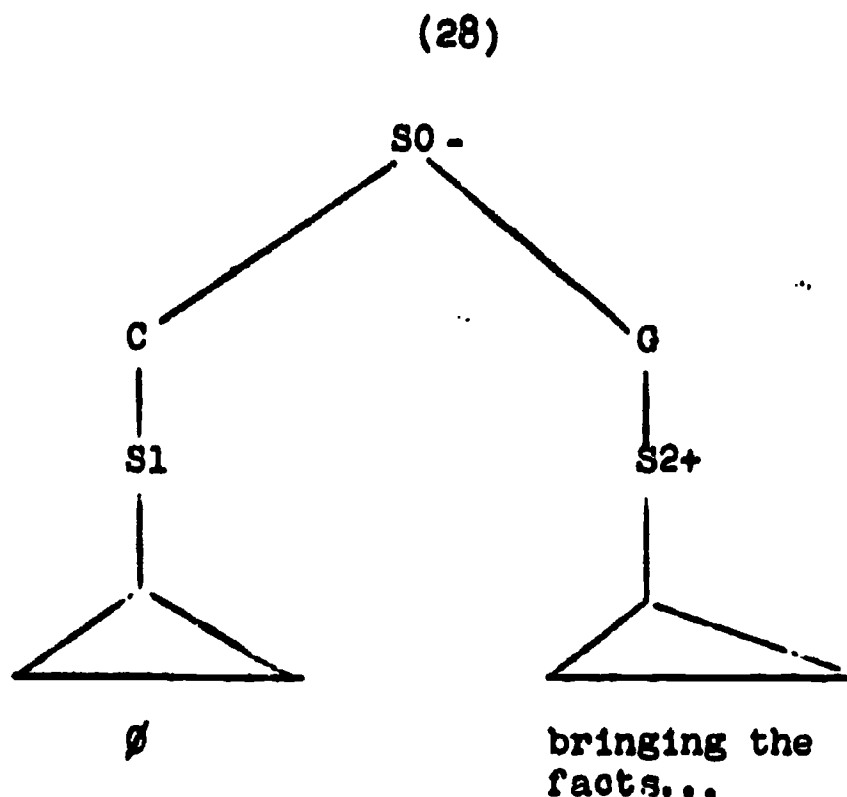
bringing the facts...

Example : (Scientific documentation -- the operation of) bringing the facts elucidated by scientific research from those who produce them to those who can make use of them (is getting out of hand) (J 97).

<u>Cp.</u> : (states the problem of) bridging the gap between the individual who wants the information and the records containing the information (J 205).

<u>Cp.</u> : (it's the intellectual job of) matching what's known and what's wanted (J 407.

<u>Cp.</u> : Finding what's known (J 407).

<u>Cp.</u> : (Discusses the problem of) getting information into the hands of the scientist who needs it (even when he doesn't know about it) (J 418).

<u>Cp.</u> : the man-document communication (problem)(J 1110).

(29)

collection of written inf.	arrangement of written inf.	distribution of written information.
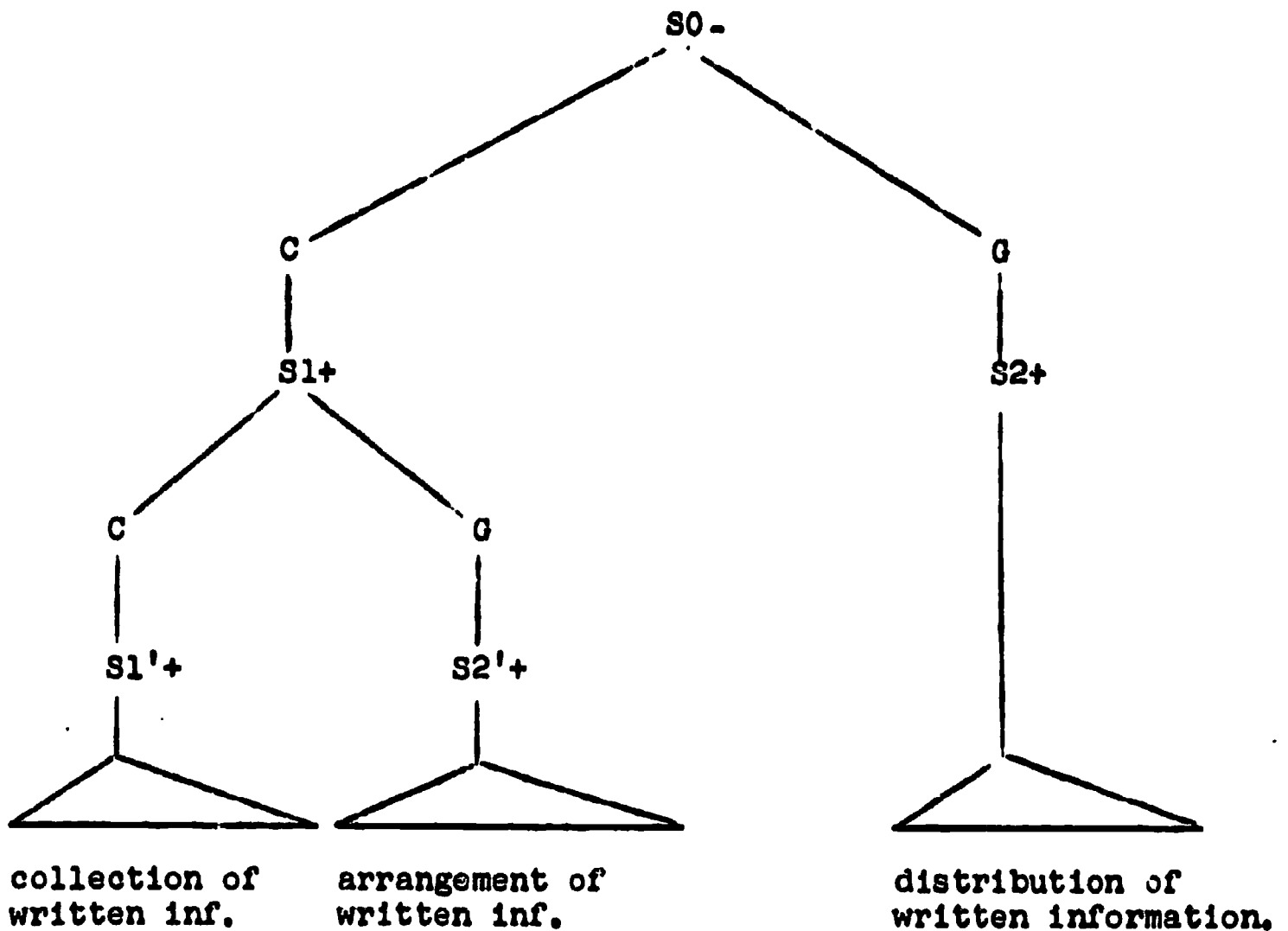
Example : (Documentation should not be contrasted
with library work it is a phase or
aspect in the chain of)collection,
arrangement, AND distribution of written
information (as found in libraries)(J 474).

(30)
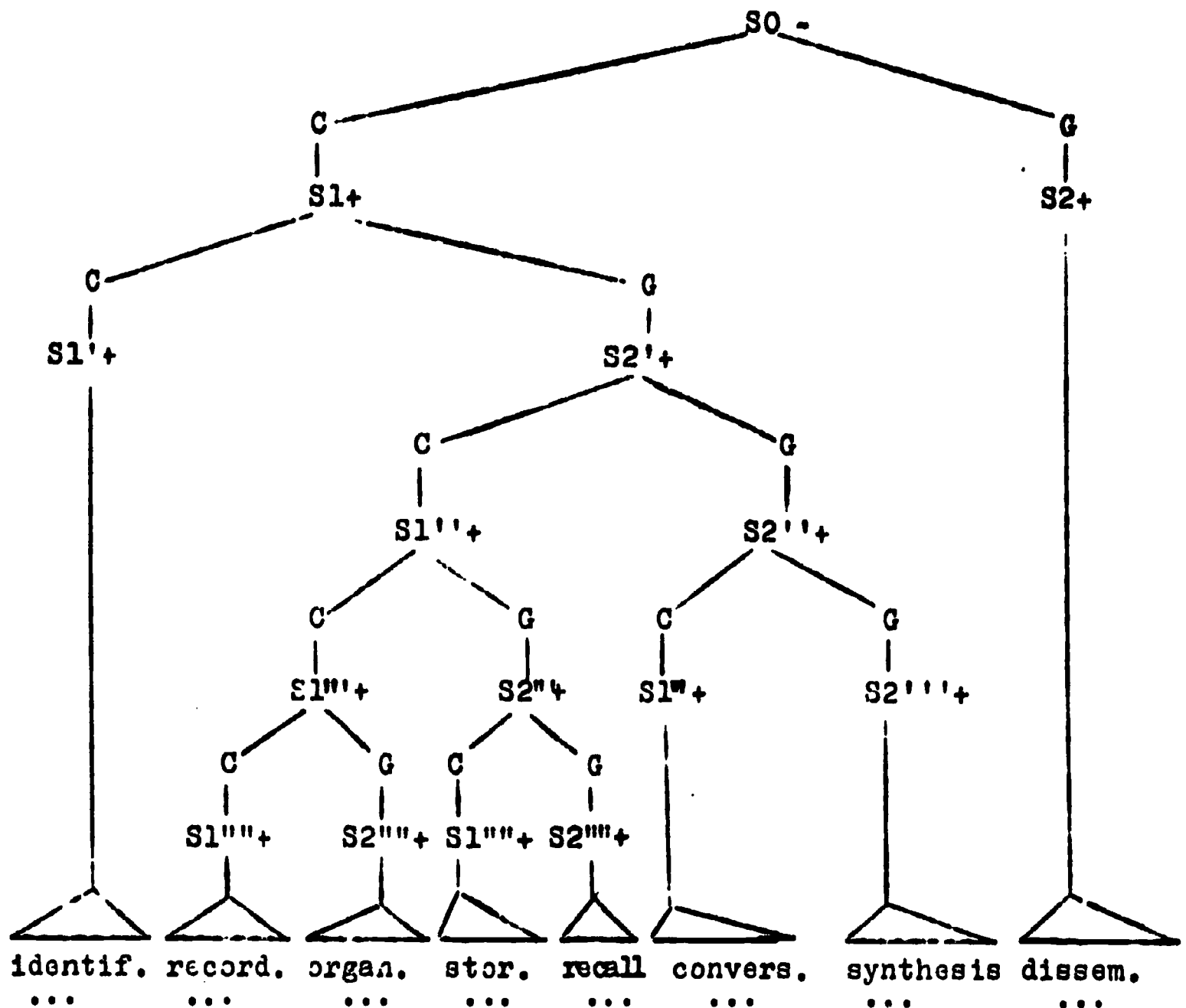
SO -

C

S1+

C   G

S1'+   S2'+

C   G

S1''+   S2''+

C   G   C   G

S1'''+   S2''+   S1'''+   S2'''+

C   G   C   G

S1''''+   S2''''+ S1''''+ S2''''+

S2+

identif.   record.   organ.   stor.   recall   convers.   synthesis   dissem.
...   .

Example : ("Documentation is best differentiated from
normal library service by the extent to which
it is concerned with a complete-cycle service
providing information. This cycle involves)
the identification, recording, organization,
storage, recall, conversion,into more useful
form, synthesis, and dissemination of the
intellectual content of print and other
recorded materials." (J 890).

(31)
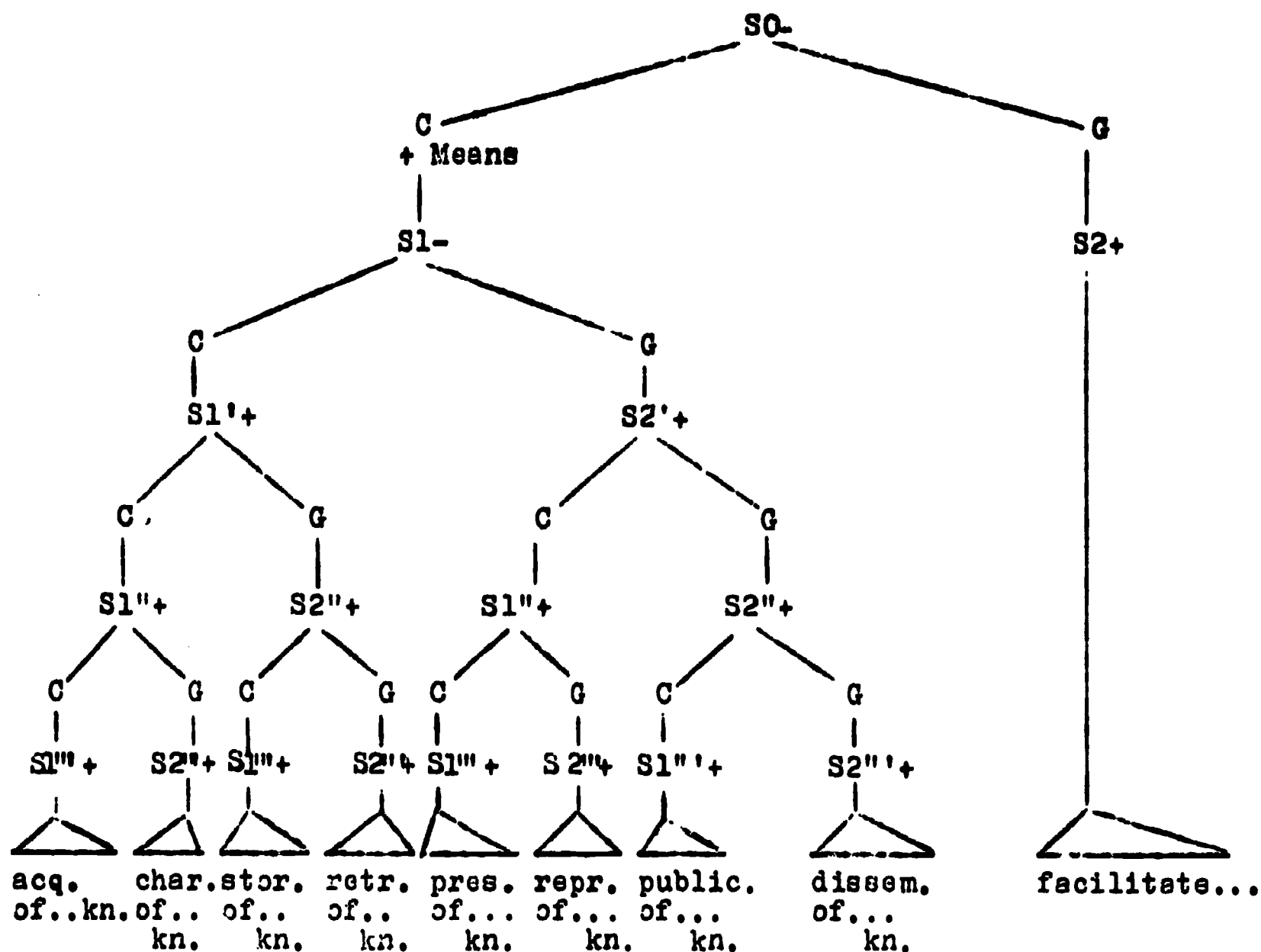
Example : (A documentation system is defined as a
complex pattern of interacting and inter-
dependent processes which) facilicates the
use of recorded specialized knowledge
THROUGH its presentation, reproduction,
publication, dissemination, acquisition,
characterization, storage and retrieval
(J 107).

(32)
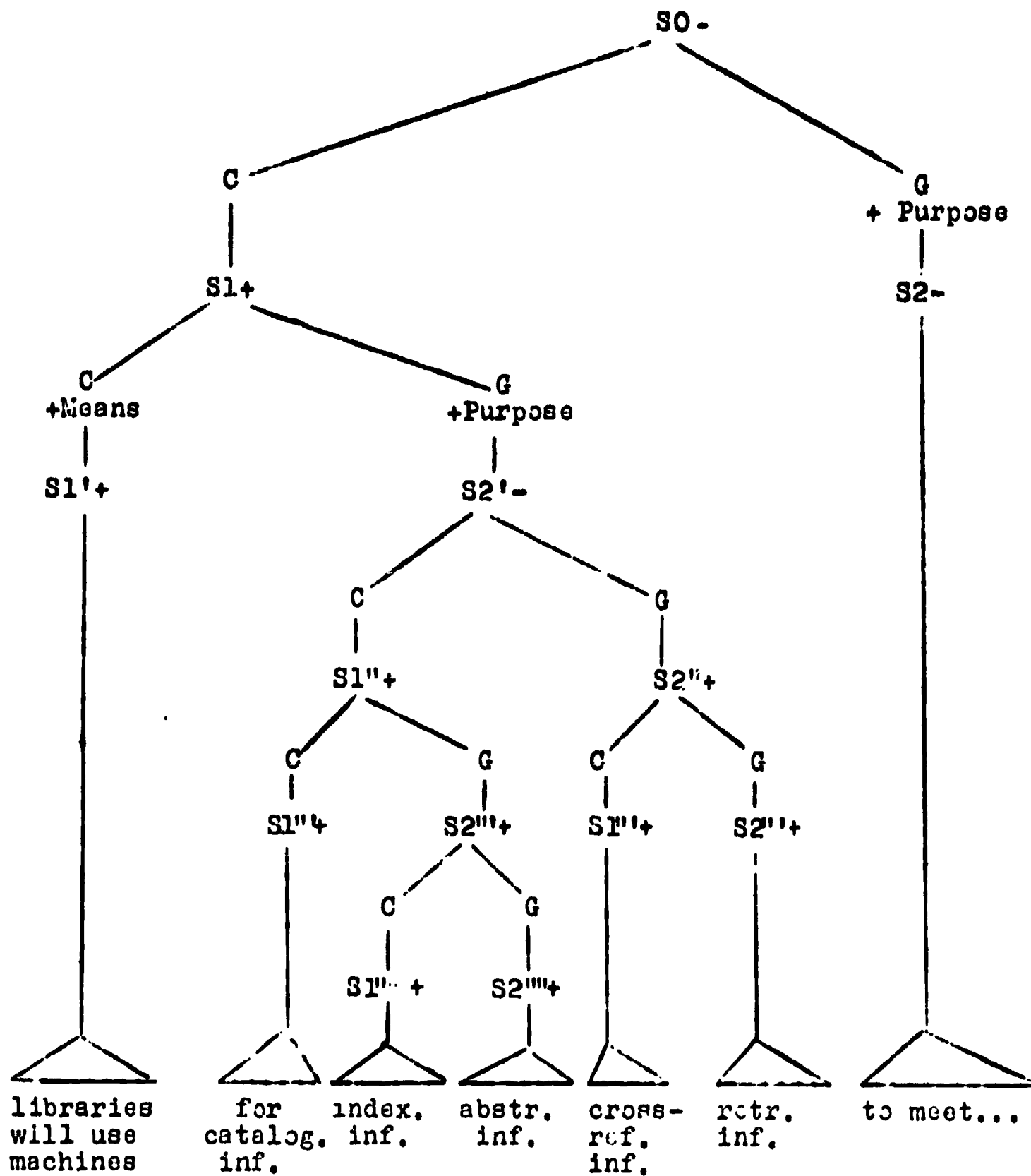
Example : TO meet the increasing value of scientific
and technical information, the libraries
will use machines FOR cataloguing,
indexing, abstracting, cross-referencing,
and retrieving of information (J 363).

(33)

Example : (The documentation problem and means
of)solving it THROUGH acquisition,
organization and storage of graphic
records, and the operation of a
retrieval system (are reviewed...) (J 497).

(34)

SO-

C
+Means

S1-

by means of
documontation

G

S2+

achieving optimum results
in communication of factual
knowledge.
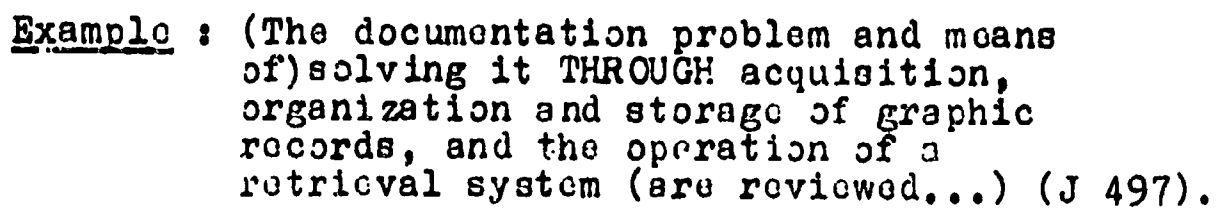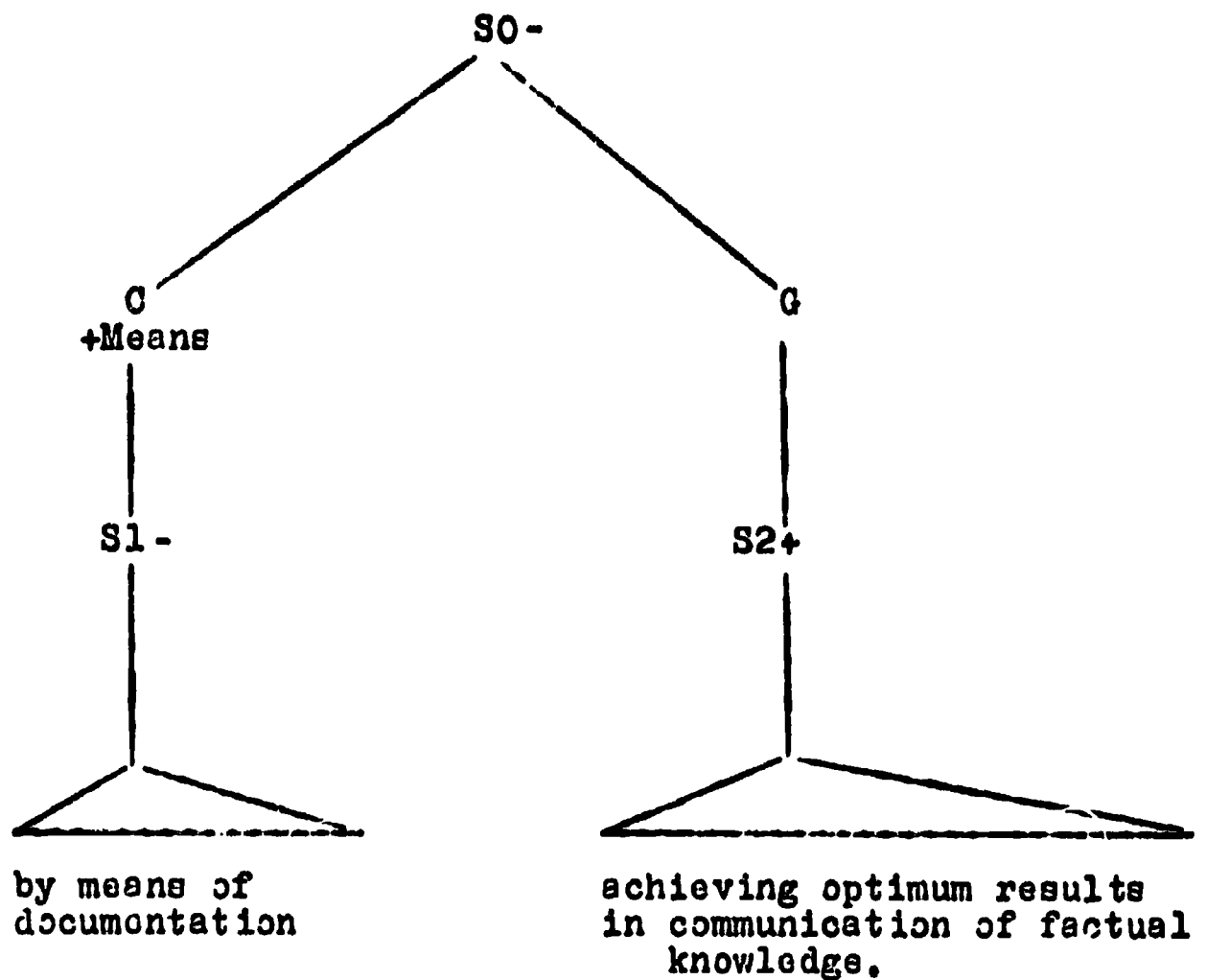
Example : (Discusses methods and techniques of)
achieving optimum results in communica-
tion of factual knowledge BY MEANS OF
documentation (J 772).

448

As a further example, I wish to quote the following abstract in order to suggest that the type of analysis I propose could be extended to discourse structures; note, in particular, the phrase <u>as a result</u> in the last sentence. I shall refrain from presenting a complete analysis of the text, however, because I do not fully understand it. I shall just present an analysis of the last two sentences in (36).

<div align="center">(35)</div>

A system is described which permits handling of abstracts in a largely mechanical way for documentation and information. The abstracts are made by using a subject index and are then recorded on magnetic tape. These subjects are coded automatically in the machine and transferred to a special "index tape" or the "index part" of the "text-in-clear-tape". The retrievals are made by coding the subjects of the question and feeding them into the machine. As a result, a list is printed containing the abstracts in question as text-in-clear (J69).

(36)
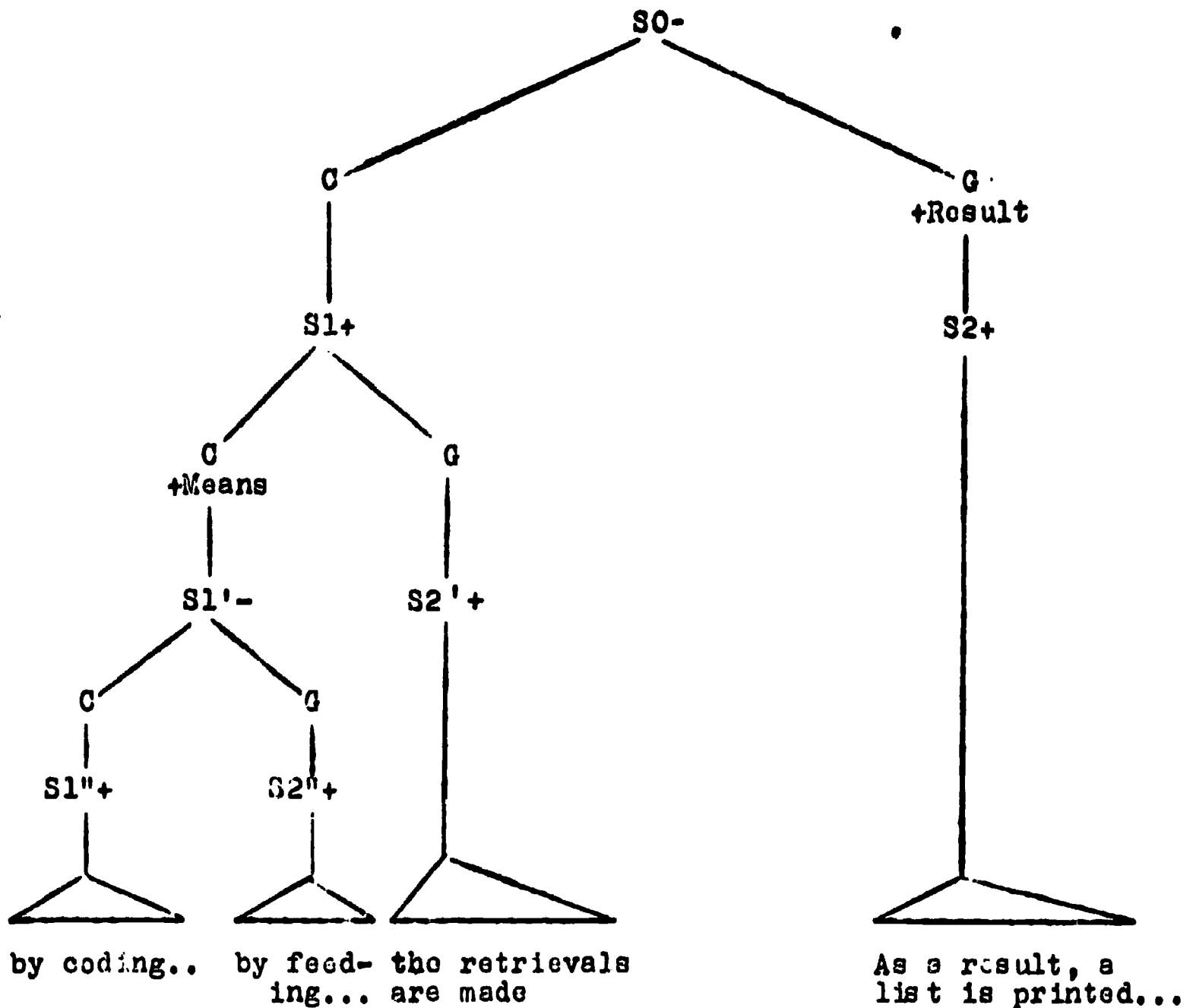
Example : The retrievals are made by coding the
subjects of the question and feeding
them into the machine. As a result,
a list is printed containing the abstracts
in question as text-in-clear (J 69).

3.4.4. A few attested examples.

The second set of analyses will also be presented in the form of diagrams. What I wish to stress here is the great variety of surface forms which can serve to express what I analyze as a Cause/Goal relation. A few features not mentioned so far will be used in the ensuing analyses. In some sentences, it is difficult to decide what (if anything) is to be the verb of the main sentence (SO). This is particularly true of modals. Thus, according to Langendoen 1970 (186-191) and many others, auxiliaries and tense markers are to be analyzed as predicates and the remaining material as their argument(s). One of the semantic arguments for such an analysis is that auxiliaries (say, 'must') can be used in the same meaning as predicates (e.g. 'necessary', or 'certain' for 'must'). I feel, however, that an alternative analysis is possible in some cases, and that it is difficult to find reasons for rejecting it. Compare, for instance (45) with (46), (52) with (53), and (68) with (69).

(37)
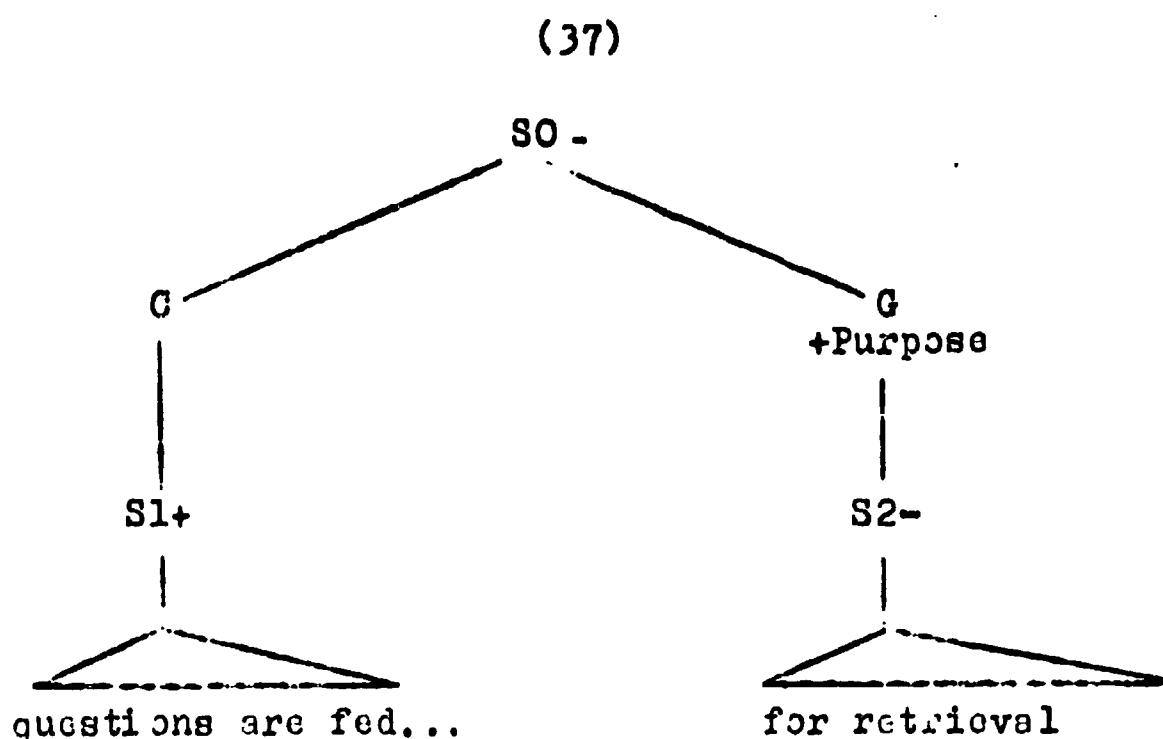
Example : FOR retrieval, questions are fed through
a standard switch board... (J 808).

(38)

SO -

C
+Means

S1-

by the direct search...

G

S2+
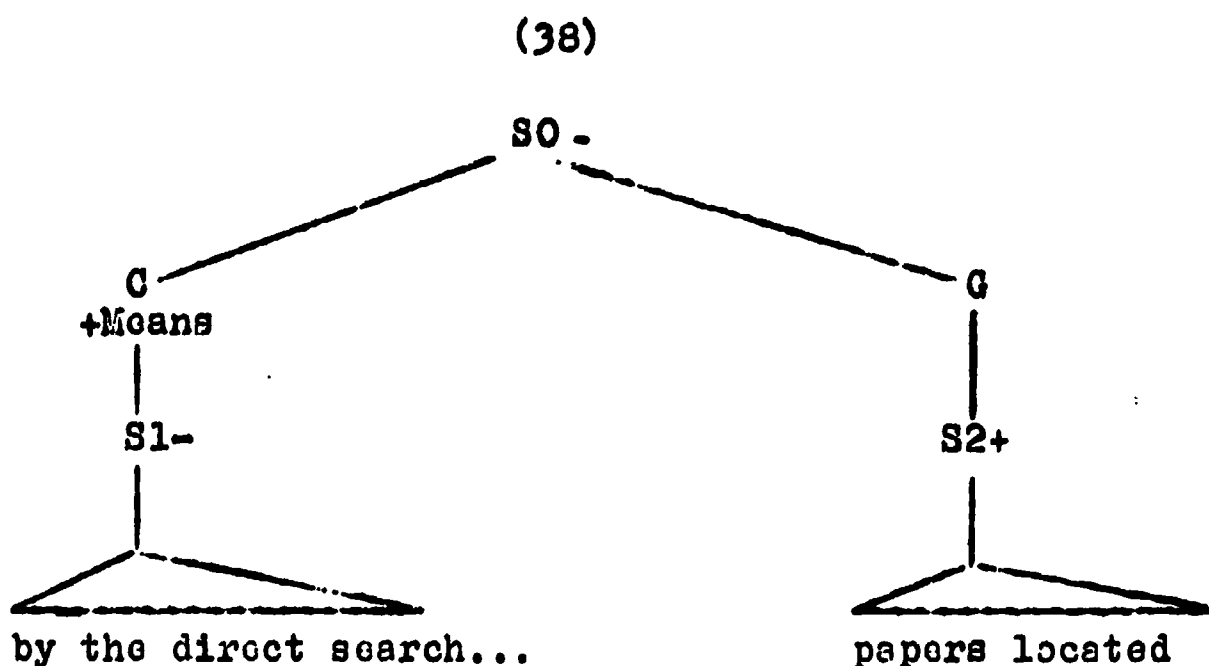
papers located

Example : papers located BY the direct search
of the abstracting journal (J 156).

(39)

SO+

V

C
+Means

S1-

G
+Purpose

S2 -

The essential of literature
purpose is... searching
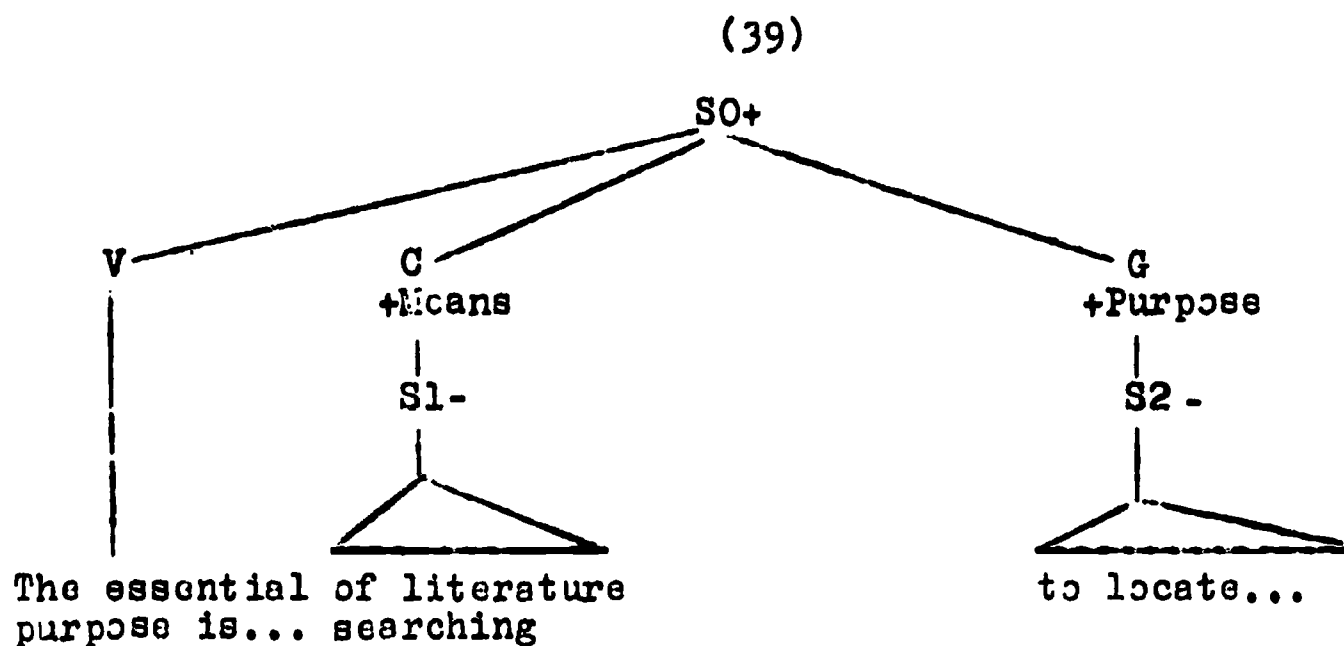
to locate...

Example : The essential purpose of literature searching
is to locate... (J 773).
Cp. : - The objective of scientific communication is
to inform (J 247).
.. The objective of the indexer is to provide...
(J 565). (= of the indexer's activity ?).
- The main purpose of this paper is to study...
(J 990). (= of writing this paper ?).
- The purpose of the procedures is to
encourage... (J 995) (= of using the
procedures ?).
- To V is the purpose of V.

444,

(40)

Tree diagram with nodes: SO at top branching to C and G; C branches to S1+, which branches to C → S1'+ ("system is operated") and G +Purpose → S2'÷ ("to identify..."); G branches to S2+, which branches to C → S1'+ ("documents are subjected to,;,") and G +Purpose → S2'÷ ("to locate...,")
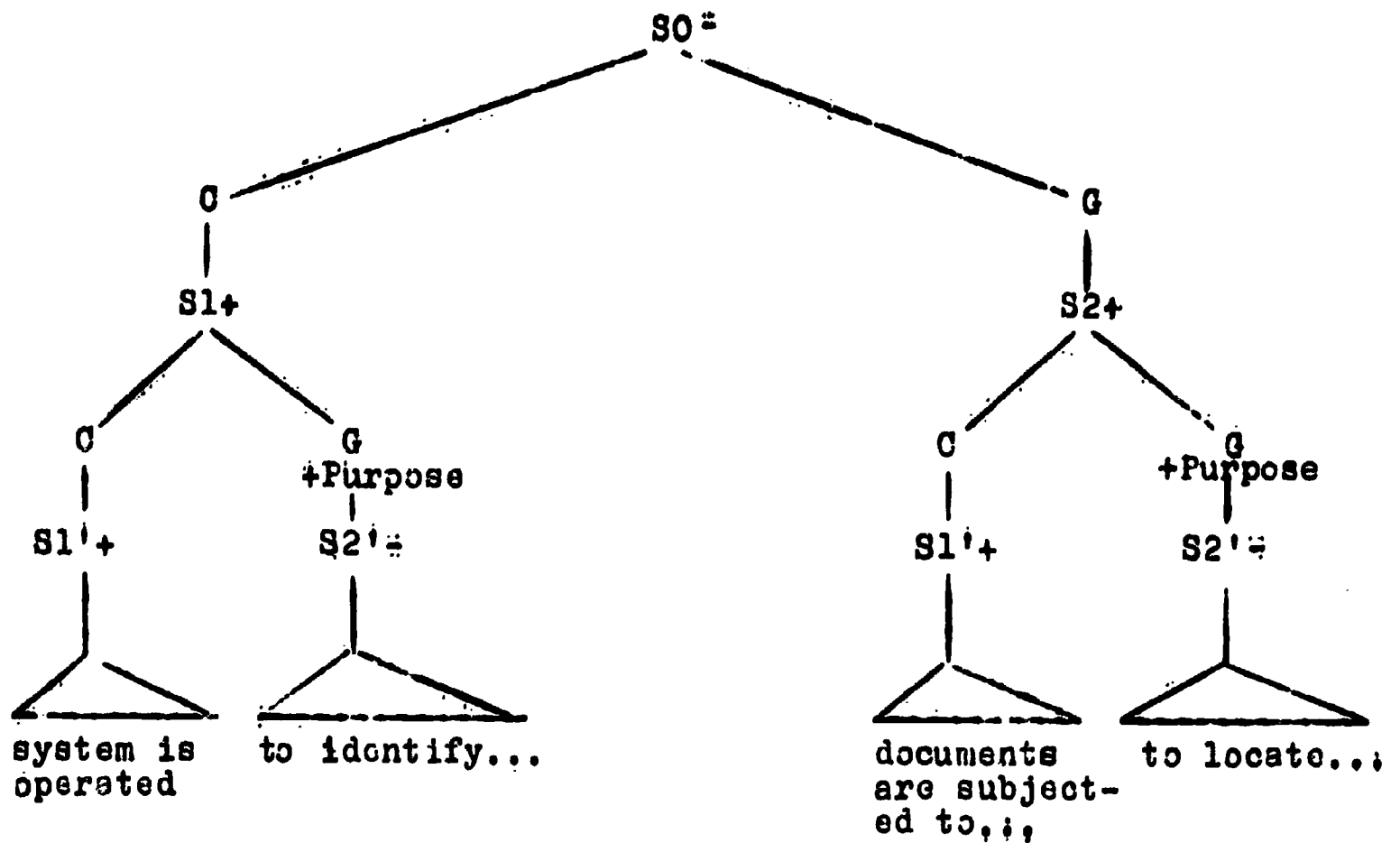
Example :   the selection system is operated TO
            identify those documents of probable
            pertinent interest; and
            the documents so selected are subjected
            to personal review TO locate the desired
            information (J 773),

(41)

```
                    SO+
           /         |         \
          V          C          G
          |          |          |
          |          S1-        S2 -
          |        /    \        |
          |       C      G       |
          |       |      |       |
          |      S1'+    S2'+    |
          |      /\      /\      /\
```

depends   on the ability   on the ability   the ability of
           of the machine   of the machine   a machine to
           to recognize...   to manipulate..   adequately
                                               determine...
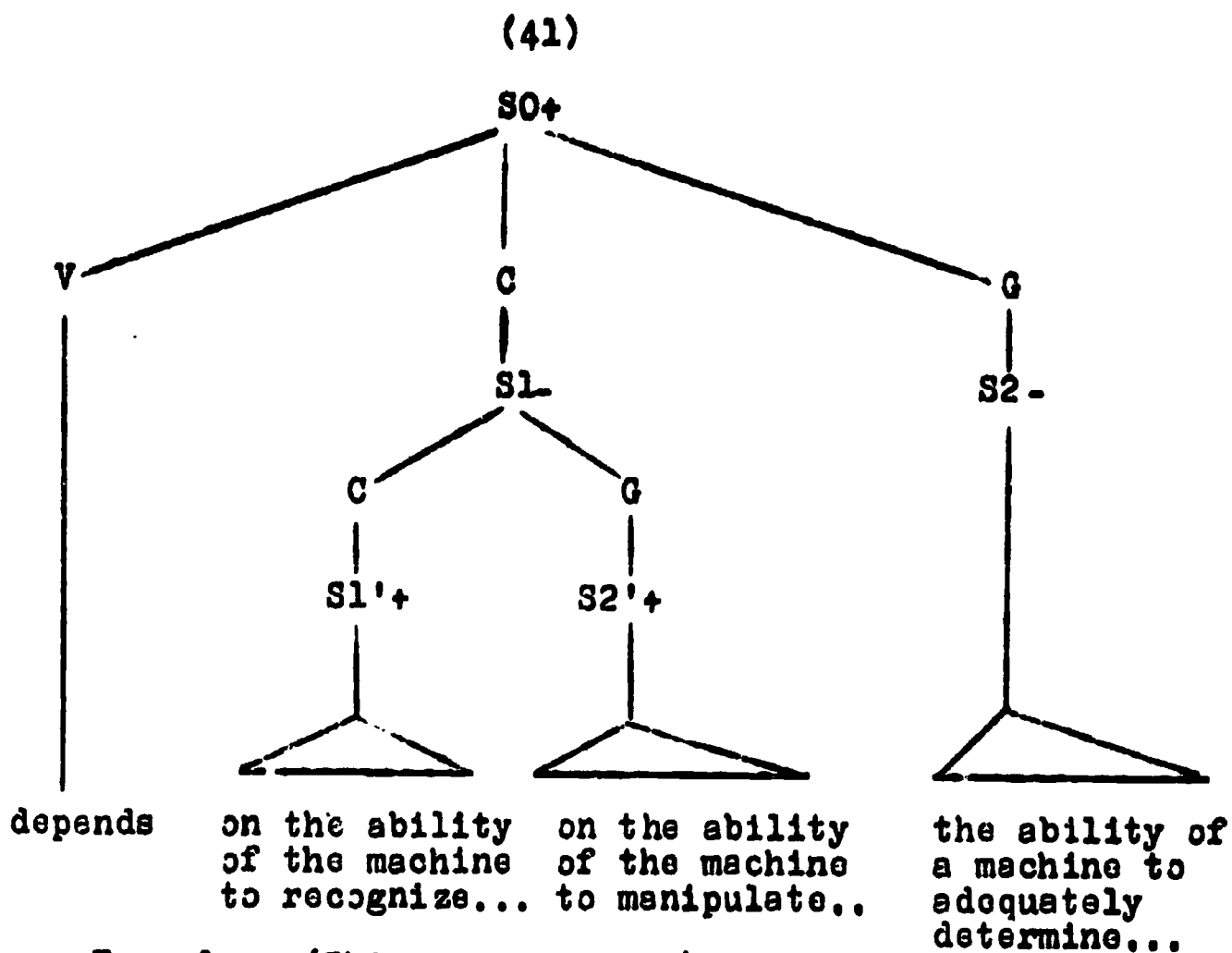
<u>Example</u> : (This asserts that) the ability of a
          machine to adequately determine information
          rele vence <u>depends</u> strongly <u>on</u> its ability
          to recognize and manipulate the syntactic
          structures of the text (: 176).

(42)

```
                    SO+
           /         |         \
          V          C          G
          |      +Time sequence  +Time sequence
          |          |              |
          |          S1-            S2 -
          |         /\              /\
```

presupposes   that it has   to got relevant
               been put in....   and valid information...

<u>Example</u> : To got relevant and valid information
          out of a machine <u>presupposes</u> that it
          HAS BEEN put in with a most exacting
          pattern of forethought (J 179).

(43)

```
                    SO+
            _____|_____
           /        |         \
          /         C          Q
         /       +Means        |
        /           |          |
       V           S1-        S2-
       |          /___\      /___\
  is possible   by searching  to receive
                one item      information...
```
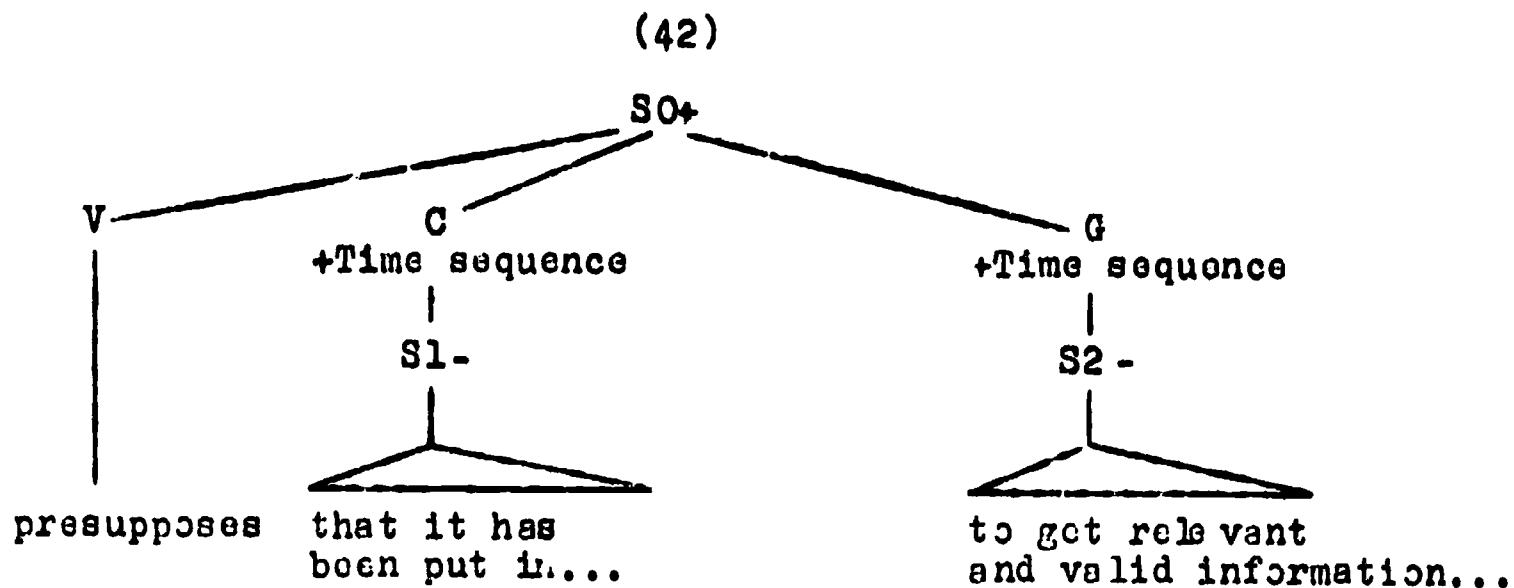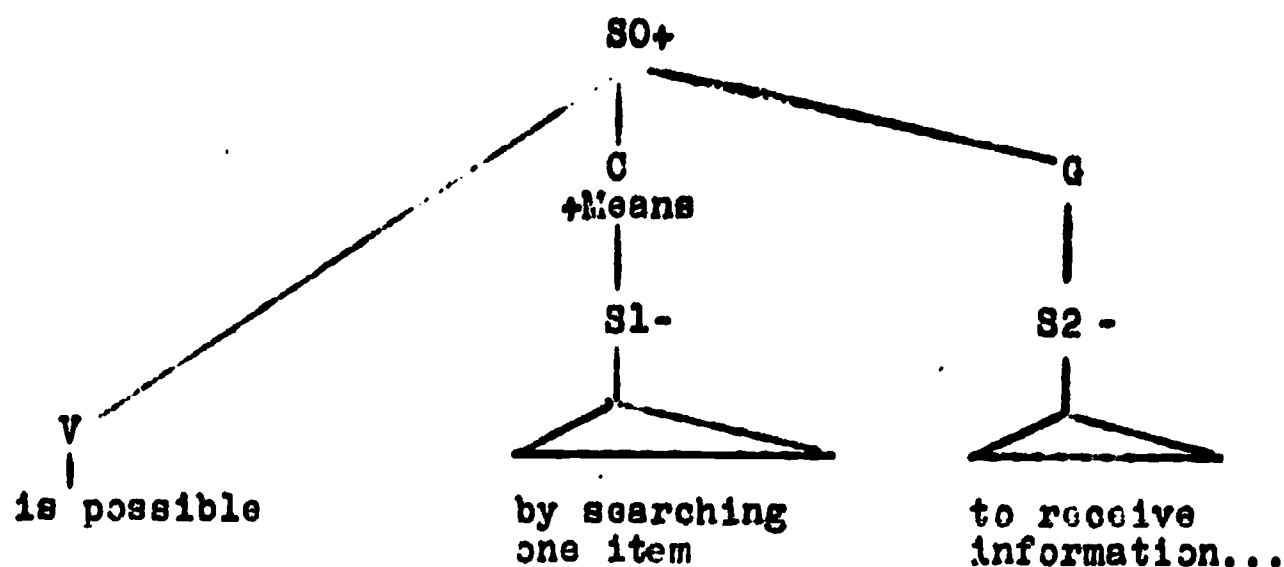
Example : ... it is possible... BY searching one item
to receive information on all compounds
of the same type (J 260).

(44)

```
                    SO+
           _____|_____
          /      ___|          \
         V  ____/   C           C
         |        |             |
         |       S1-           S2-
         |      /___\         /___\
    is unable  selection...   to produce the
                              retrieval
                              selective results...
```
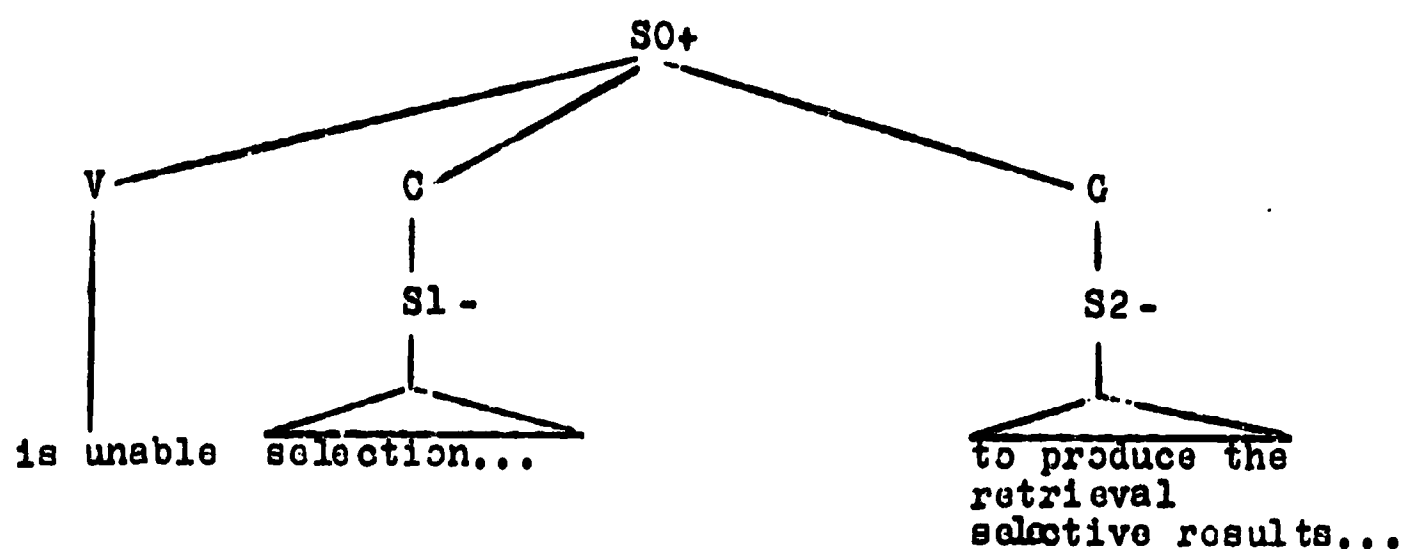
Example : (It is shown that) selection according to
"NOT" is unable to produce the retrieval
selective results ordinarily claimed for
this operation... (J 680).

(45)

```
                        SO+
                   .─────┼─────────────
                  /      │              \
                 /       C               G
                /    +Condition          │
               V                         │
               │        S1_             S2_
               │       ╱───╲          ╱────╲
            Could    if requesters...  searches become...
```
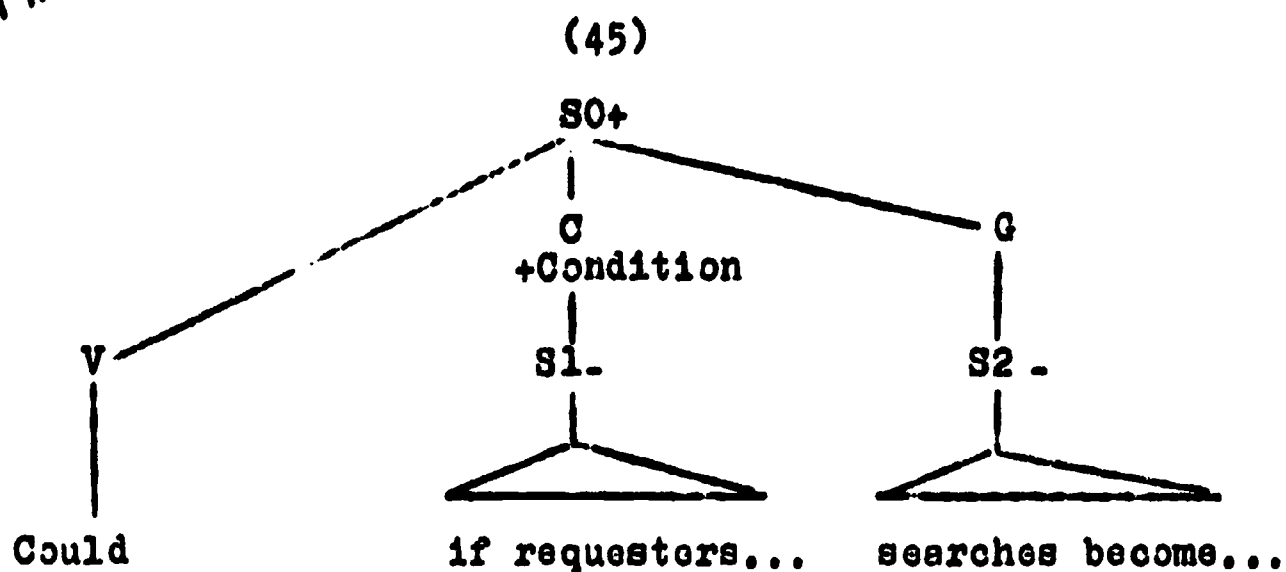
Example : Searches could become more effective if
requesters designated a plurality of pertinent
subject headings arranged according to
importance (J 266).

(46)

Alternative analysis

```
                        SO-
                   ─────────────────────
                  /                      \
                 C                        G
             +Condition                   │
                 │                        │
                S1-                      S2+
              ╱────╲                   ╱────╲
          if requesters...         searches could become...
```

(47)

SO+

V   C.   G
        +Condition

        S1-   S2-

is possible only if...   the change over...
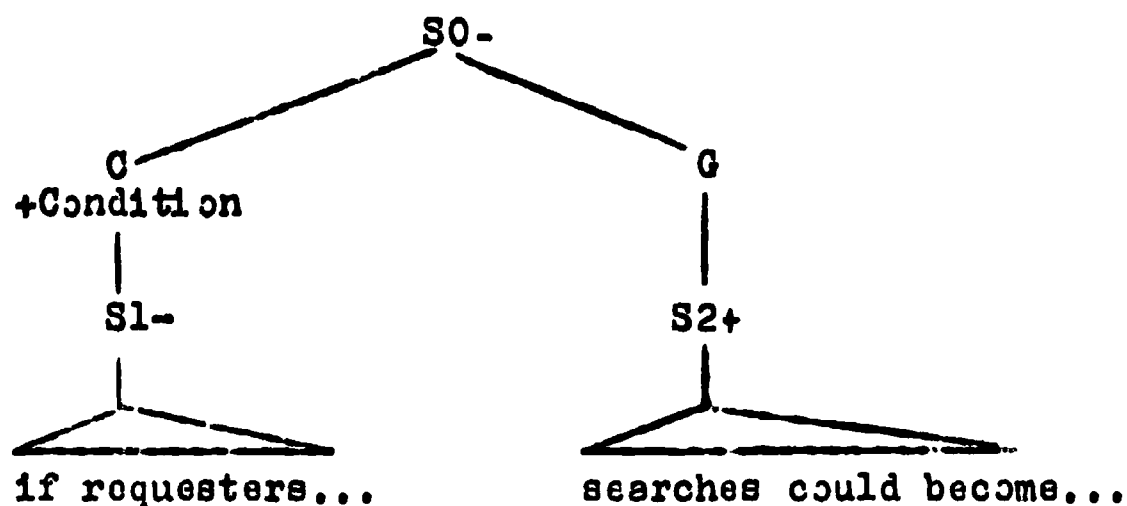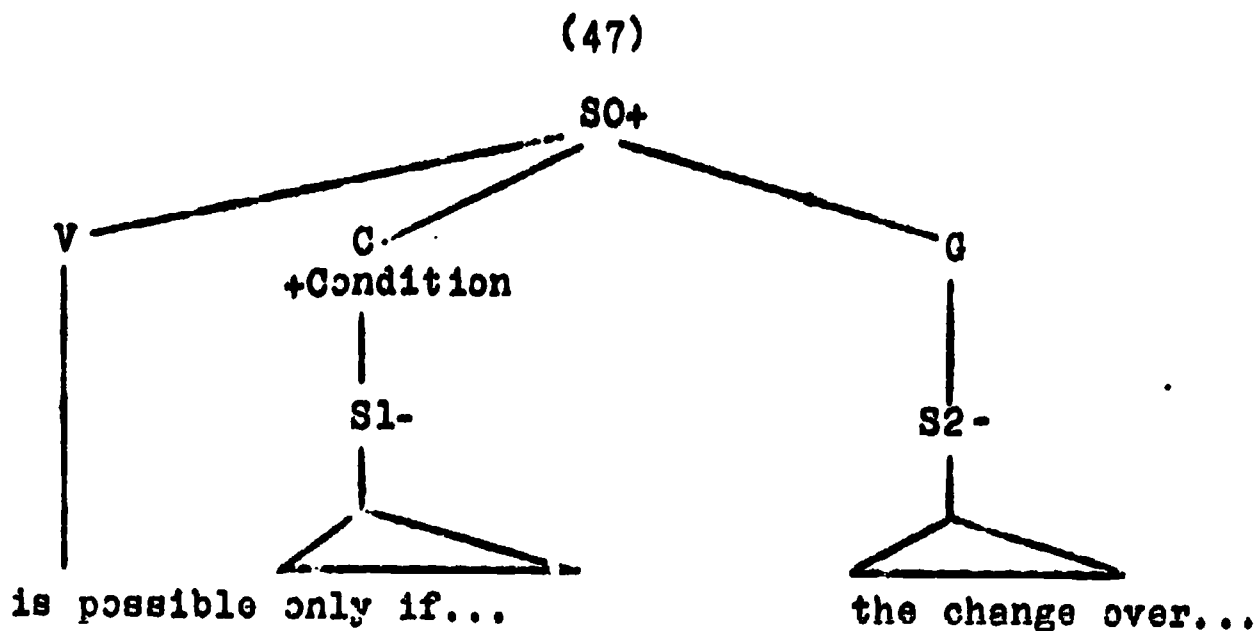
Example : (In the system) the changeover to
mechanical selection is ONLY possible
IF a previously defined list of concepts
is available (J 537).

(48)

SO -

C   G
+Time sequence
+Condition

S1-   S2+

                              C   G

                              S1'+   S2'+

when...super-   light passes...   light indicates...
imposed
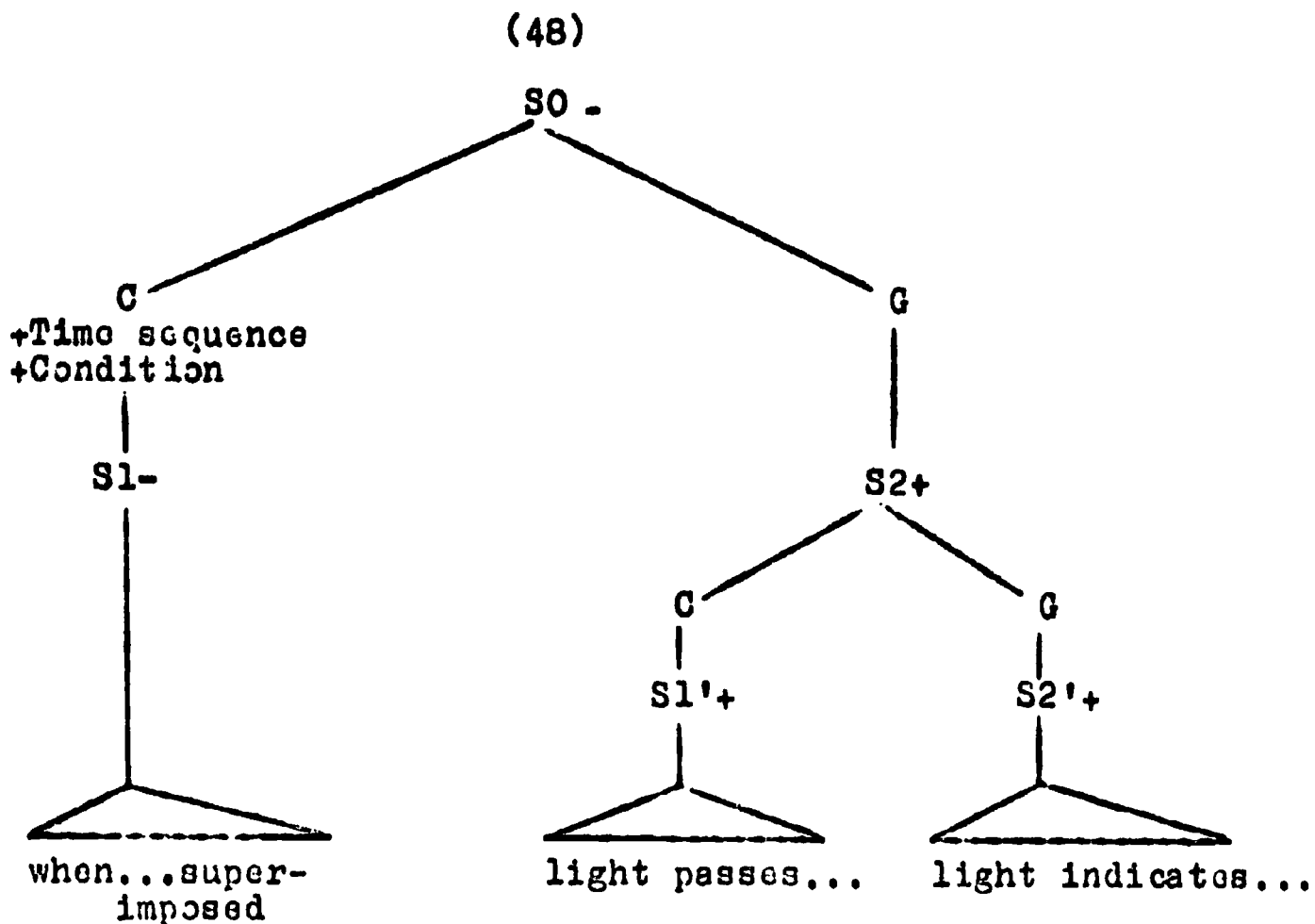
Example : WHEN several subject cards are superimposed,
light passes through the holes in the subject
cards in that particular pack and indicates
which documents pertain to all of those
subjects (J 1099).

(49)

SO -

C
+Condition

S1-

G

S2+

C

S1'+

G

S2'-

Onco the subject
heading list has
becn set up

highly trained
specialists
aro not necded

to opdrate
the system
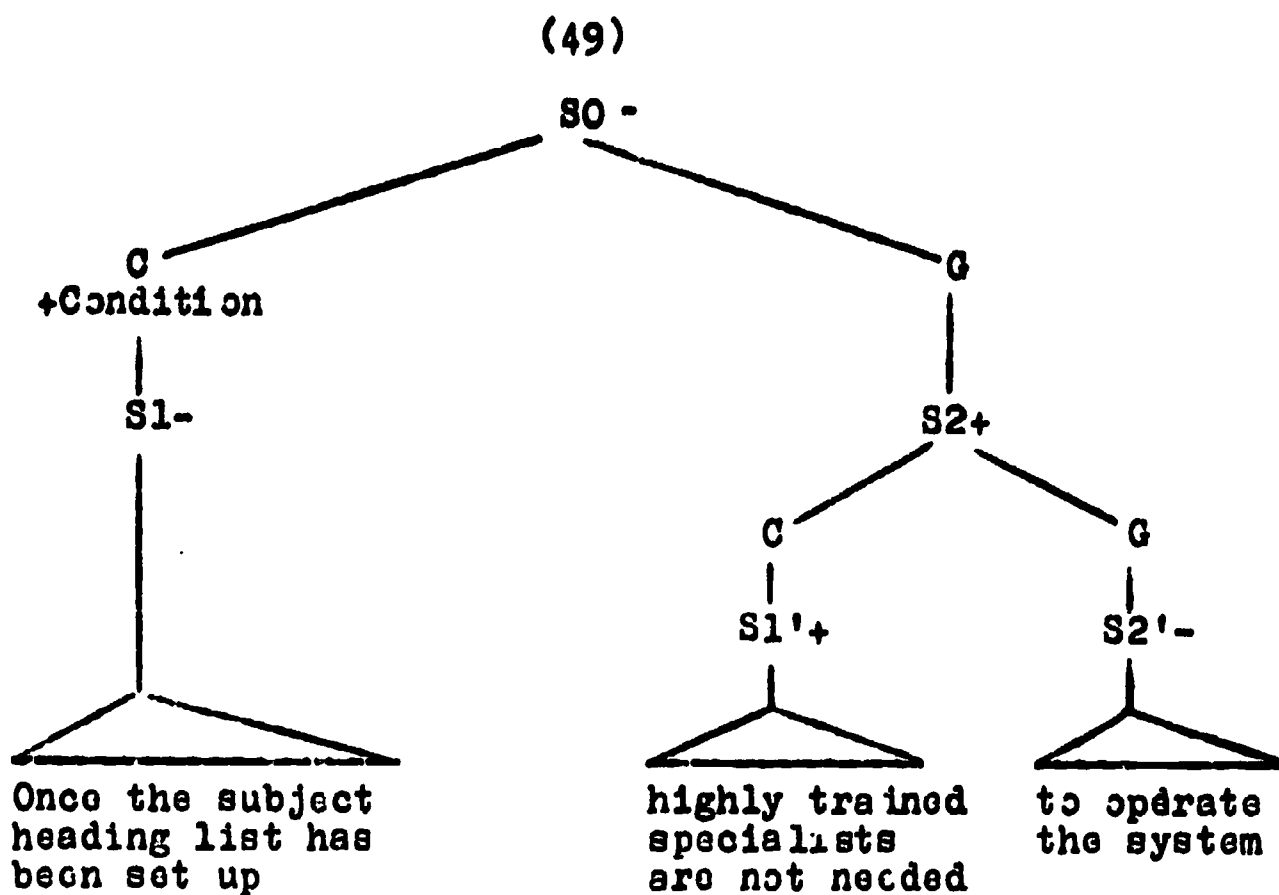
Example : ONCE the subject hoading list has boon set
up, highly traind specialists are not needed
TO oporate the system (J 843).

(50)

SO -

C

S1-

G

S2+

without disturbing
indoxing alrcady
done

the subject hoading
list can bo enlarged
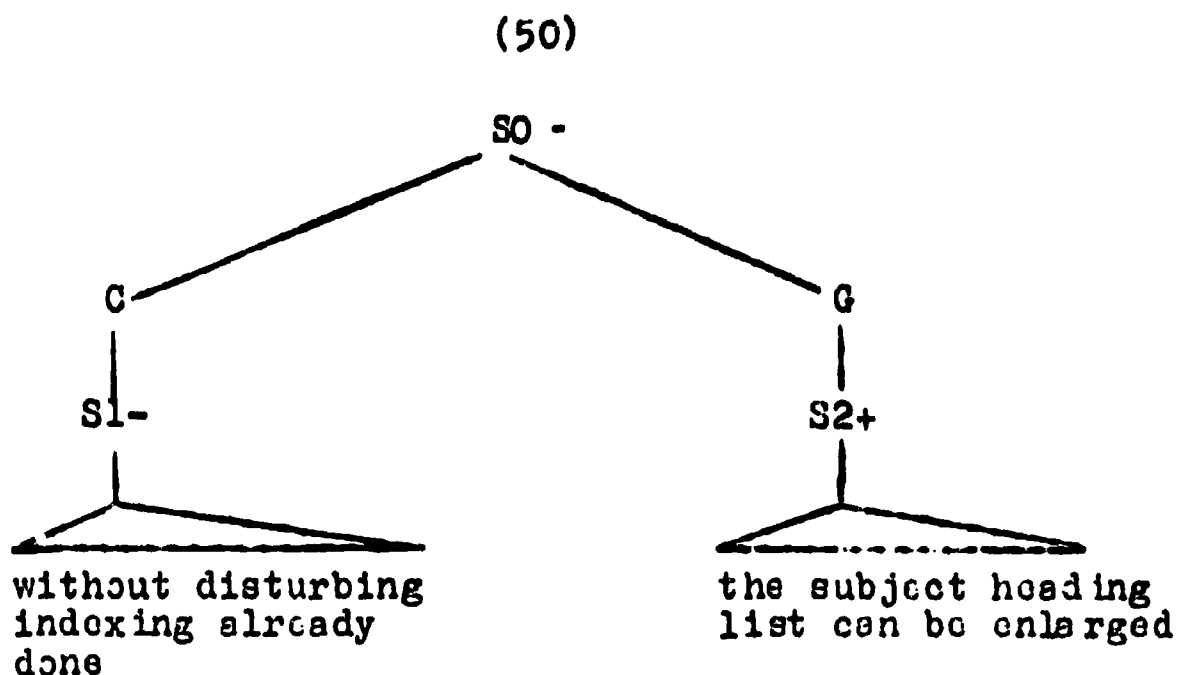
Example : The subject hoading list can be enlarged
WITHOUT disturbing indcxing alrcady
dono (J 843).

(51)

```
                    SO+
                  /  |  \
                 /   |    \
                /    C      G
               /     |      +Time sequence
              /      |           |
             V      S1-          S2-
             |      /\           /\
             |     /  \         /  \
        is necessary  to develop...   before any automatic
                                       system ... can be
                                       effective
```
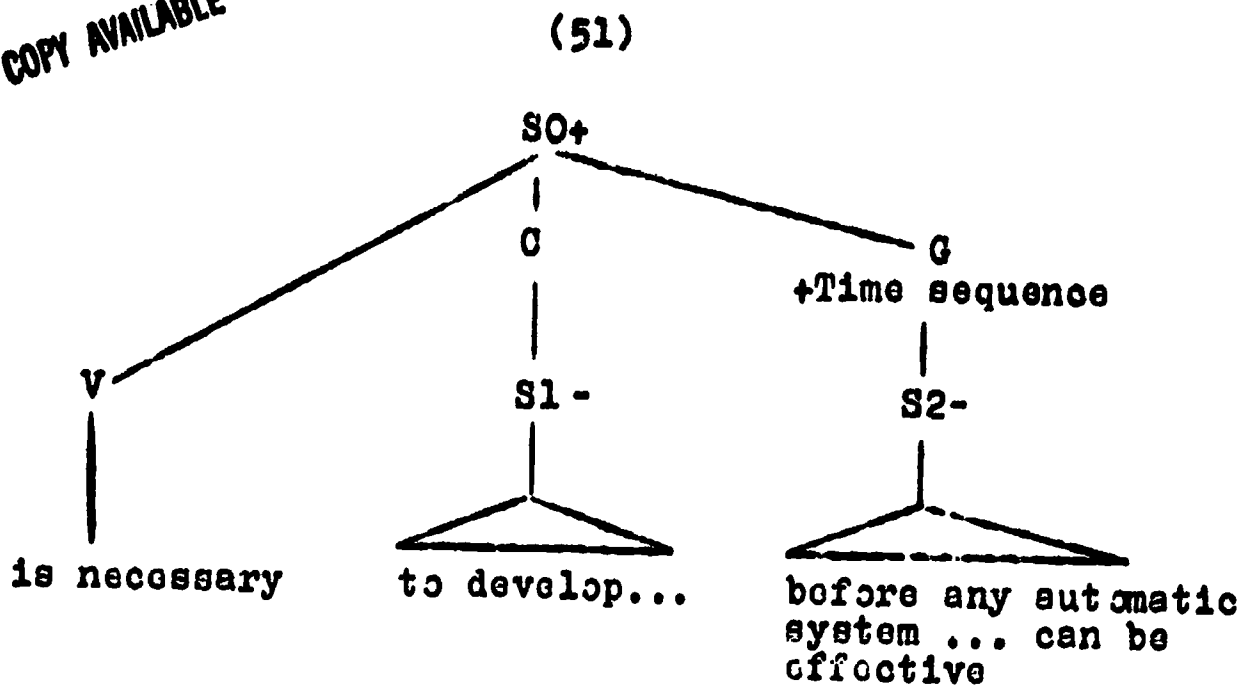
Example : BEFORE any automatic system for storing,
searching, and retrieval of information
can be effective, it is necessary to develop
an efficient and economic method of placing
and locating the information in the system
(J 404).

(52)

```
                    SO+
                  /  |  \
                 /   |    \
                /    C      G
               /     |      |
              V     S1-     S2-
              |     /\      /\
              |    /  \    /  \
            must  a researcher find. if a researcher is to...
```
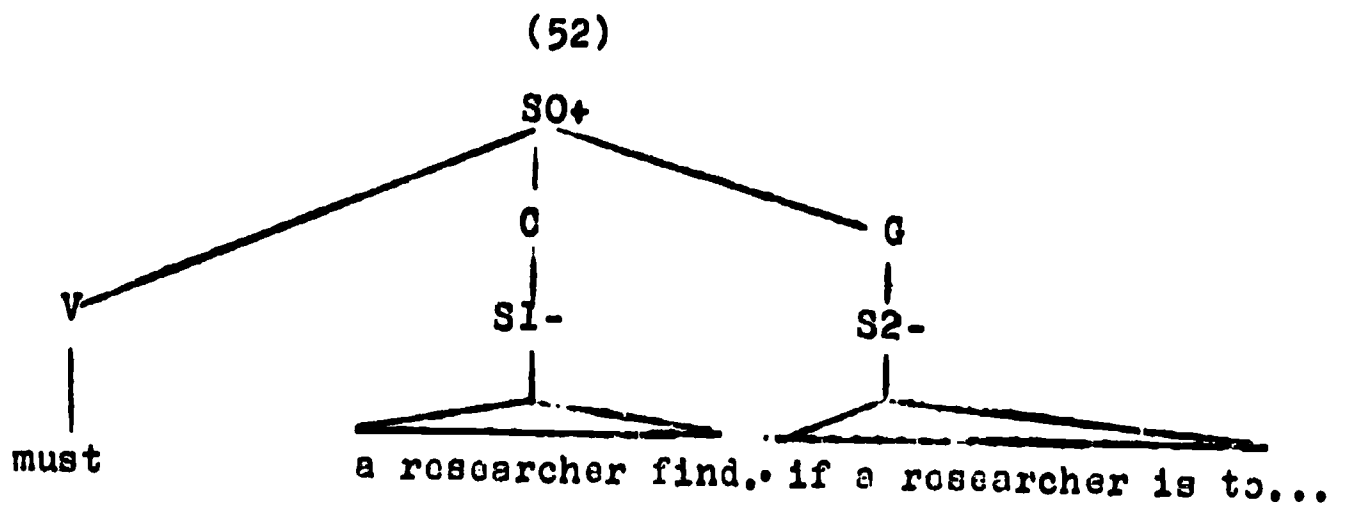
Example : IF a researcher is to avoid duplicating
work already done somewhere else, he must
find what pertinent work exists anywhere...
(J 637).

(53)

Alternative analysis

```
                    SO -
                 /        \
                C           G
                |           |
              S1+          S2-
             /‾‾\         /‾‾\
```

A researcher must   if a researcher
find...   is to avoid...

(54)

```
                    SO -
                 /        \
                C           G
           +Condition       |
                |          S2+
              S1-         /‾‾‾\
            /‾‾‾\
```

unless it can handle   no large scale information
searching for complex   retrieval system can be
relationships   a success
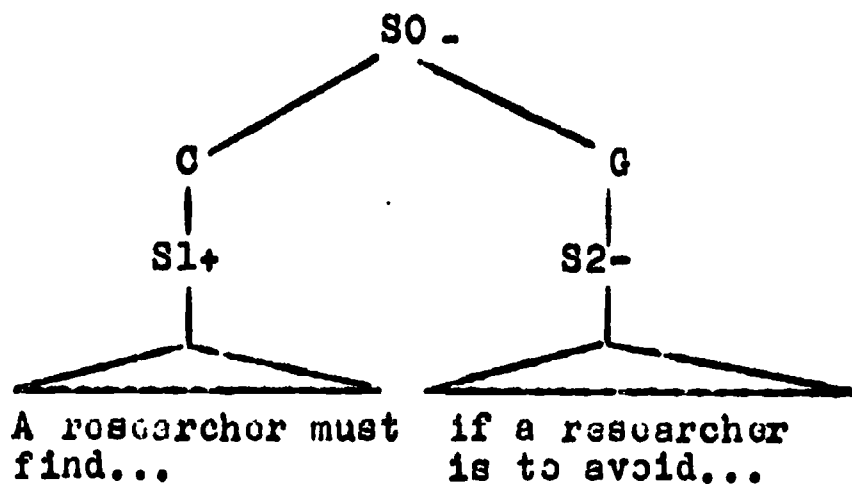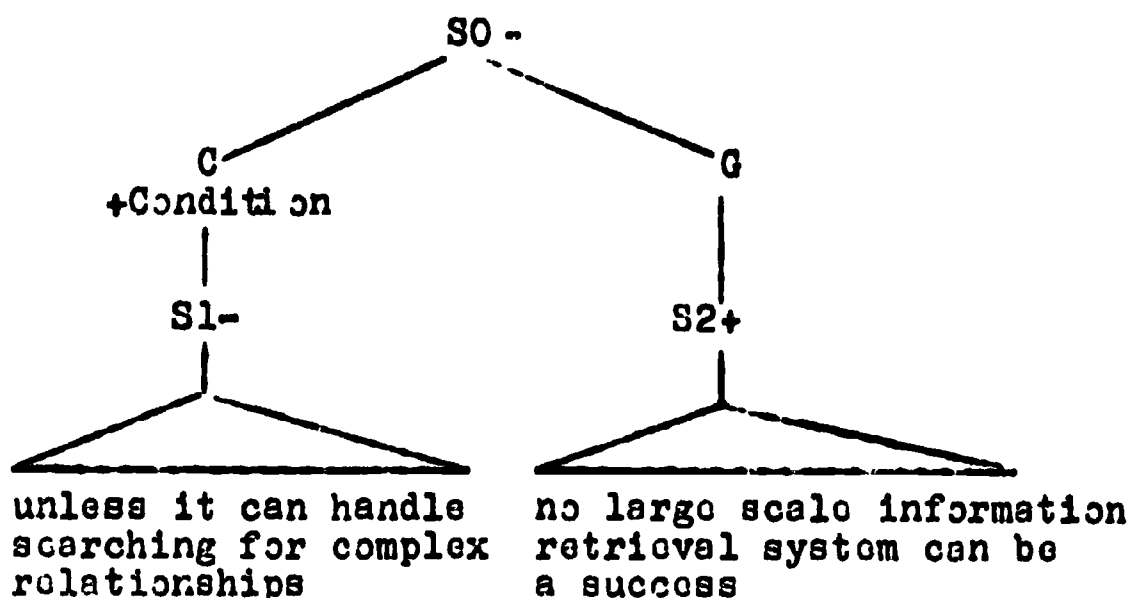
Example : (The writer feels that) no large scale
           information retrieval system can be
           a success UNLESS it can handle searching
           for complex relationships (J 20).

(55)

```
                    SO ?
            /        |        \
           V         C          G
           |         |          |
           |        S1         S2
           |       /‾‾\       /‾‾\
```

as a conse-   the...expansion..   the published literature
quence of   to incr.ase...

Example : The published literature to incr.ase in quanti-
           ty and complexity as a consequence of the
           expected continuous expansion of research
           activity. (J496).

(56)

```
                        SO -
              C                      G
              |                      |
            S1-                     S2+
         +Condition
        C          G                 |
     +Reason      S2+
        |          |                 |
       S1-
    either becauso..  when...   publication is
    or becauso...                  useless
```

Example : Publication is useless WHEN it neither
informs nor ontertains, WHEN it solves no
old problem or stimulatcs no now rosoarch,
VHEN it serves no reference function,
oither BECAUSE tho theme is trivial or
tho information is alroady available and
accessiblo in the lit_raturo (J 665).

(57)

```
                                    O+
     V                  C                       G
                     +Reason
                        |                        |
                       S1+                      S2-
              V        C          G               |
                    +Means
                      S1'-       S2-
    noces-  pos-   by utiliz-  to doscribo...   to dovolop...
    sary    sible  ing...
```

Example : BECAUSE it is possible to dorc_ibo ono
concopt BY utilizing sovral different verbal
exprossions, it (was found) nocessary to
dovelop ono metalinguistic common danominator
for a mechanized information rctrieval system
(J 707).

(58)

```
                    SO_
                  /      \
                 /        \
                C          G
              +Reason      |
                |          |
               S1-         S2+
              /|\         /|\
        because questions    this simplifies soarch
        aro taken...
```

Example : (At... documents are indoxod by both UDC
and Unitorms). This (...) simplifios search
BECAUSE questions are takon diroctly to the
most appropriate indox (J 872).

(59)

```
                    SO_
                  /      \
                 /        \
                C          G
                |          |
               S1-         S2+
              /|\         /|\
        bocauso of tho      furthor dovclopmonts
        massivo incroasos...   ... will tond...
```

Example : (Prodicts that) furthor dovolopmonts in
documontation will tond to bo in tho
diroction of olectronic storago units
BECAUSE OF tho massive incroasos in the
matcrial to be handlod (J 785).

(60)

SO -

C
+Condition

S1-

oven though
documents may
not bo...

G

S2+

system can find
documents...

Example : (Describes an all-computer document-
retrieval system) which can find
documents related to a request EVEN THOUGH
they may not be indexed by the exact terms
of the request... (J 942).

(61)

SO -

C
+Means

S1-

C

S1'+

By noting...

G

S2'+

with due regard...

G

S2+

C

S1'+

documents are
selected...

G

S2'+

documents are
arranged...
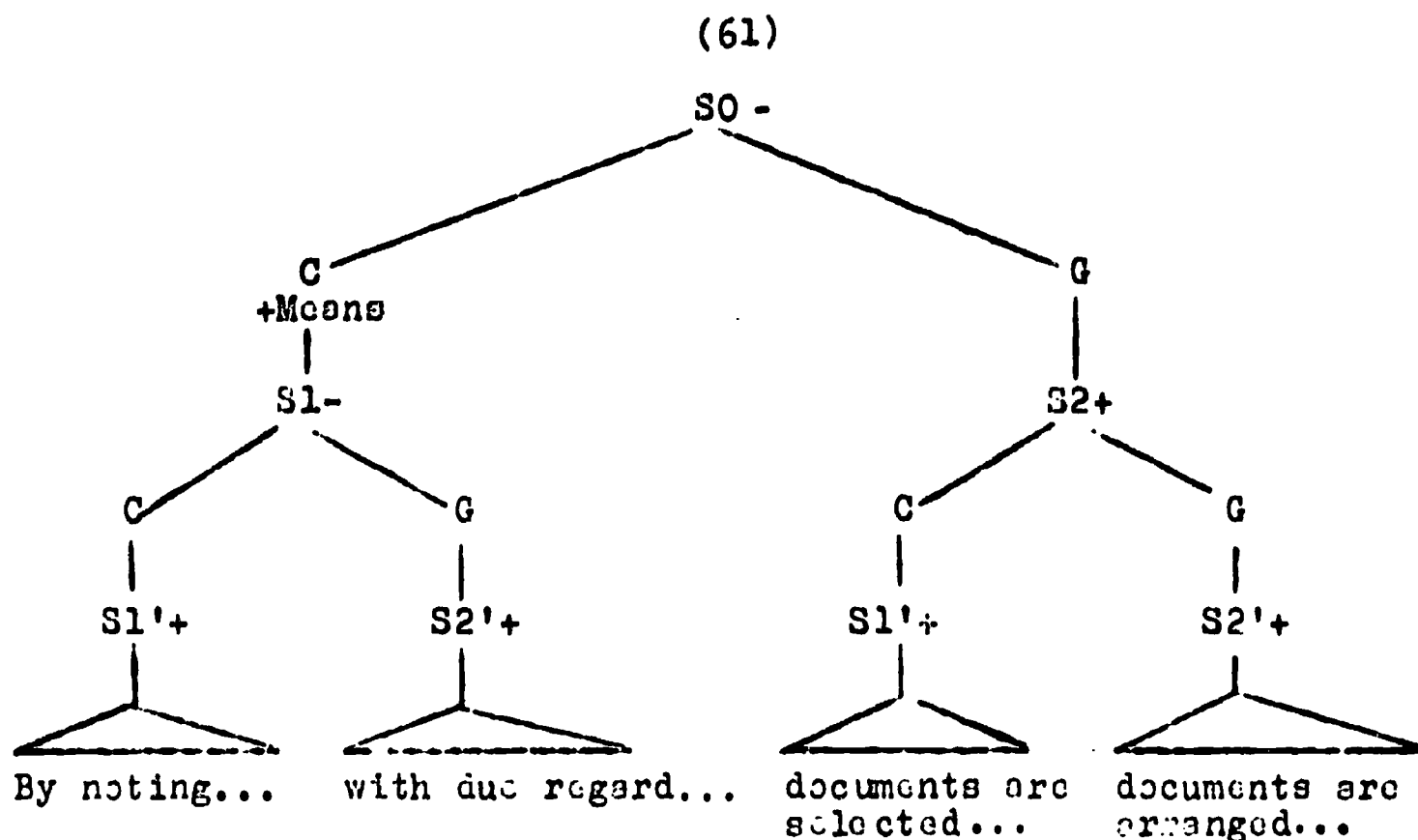
Example : BY noting the number of matching terms
between this extended list of request
terms and the terms used in indexing a
document, and WITH due regard for their
degree of association, documents are selected
by the computer and arranged in the order
of their relevance to the request (J 942).

(62)
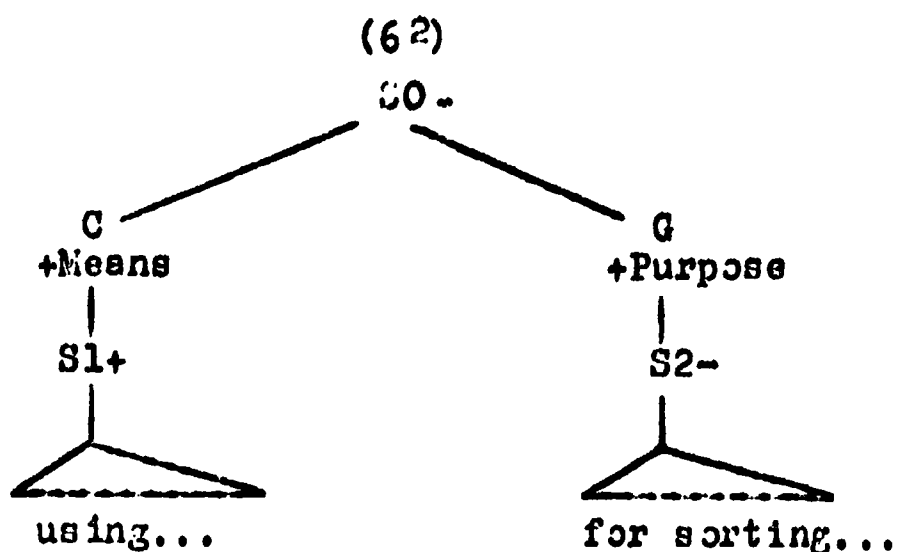
Example : (This proposes) using a high-speed,
sequence-controlled digital computer
coupled to punched card reading and sorting
equipment, called Electronic Structural
Correlator (ESC) FOR sorting all chemical
compounds containing a specified chemical
group (J 231).

(63)
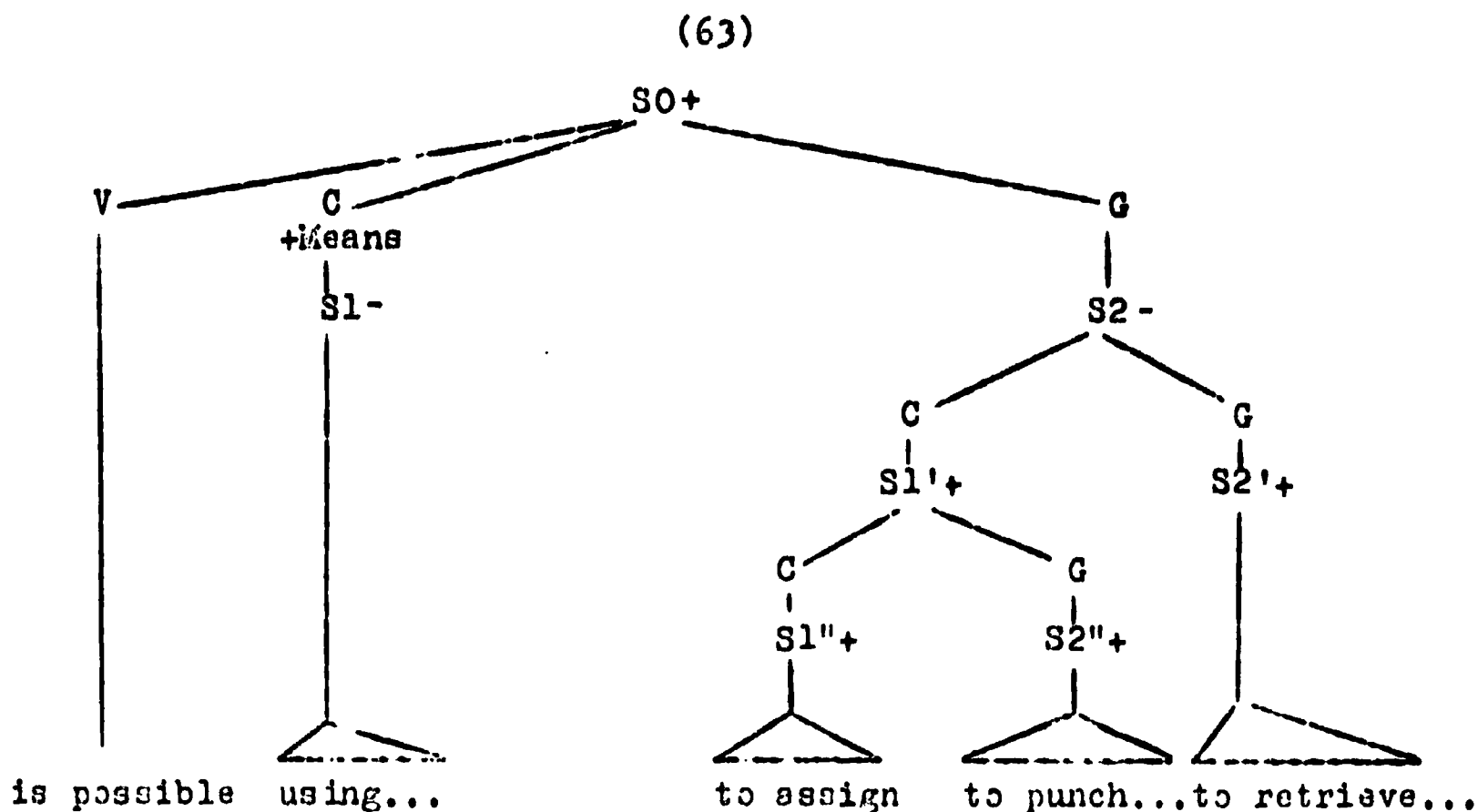
Example: Using the officially accepted International
System of Notation of Organic Compounds of the
International Union of Pure and Applied Chemis-
try, a modified IBM 026 Card Punch with document
writing feature and an IBM 466 electric type-
writer, it is now possible to assign to each
organic compound a single mathematically unique
group of descriptors, AND to punch these into a
card AND retrieve the original cipher in type-
written form at the rate of 500 characters per
minute (J 273).

(64)


```
                        SO -
                 /              \
                C                G
                |                |
               S1+              S2+
              /    \           +Purpose
             C      G         /    |    \
          +Reason   |        V     C     G
             |      |        |    S1'-   |
            S1'-   S2'+       |     |    S2'-
            / \    / \       / \   Ø=S1  / \
           /___\  /___\     /___\  |    /___\
        Based on... ...conducted  The pur     to determine...
                              pose was
```
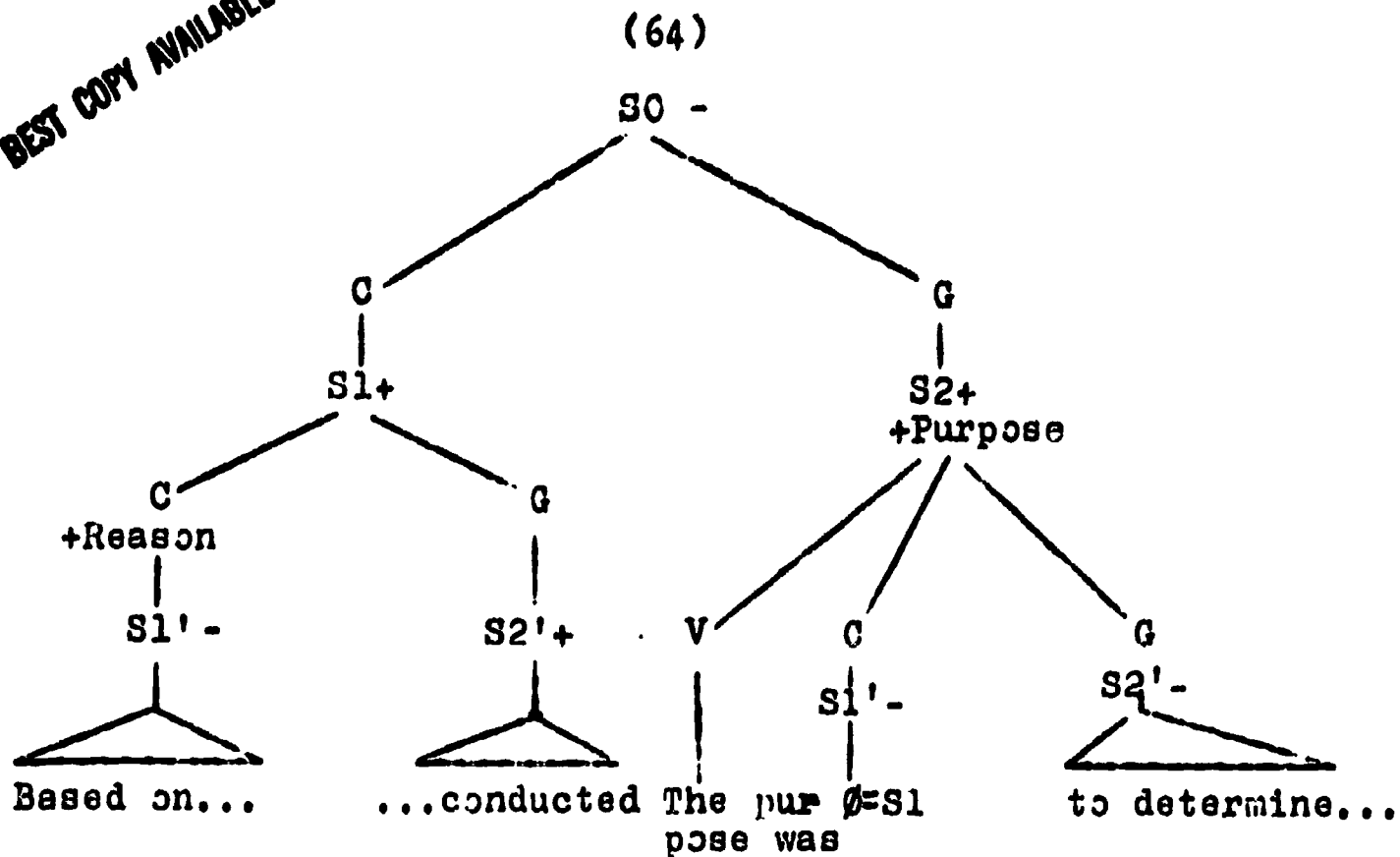
Example : Based on the growing interest in cooperative
effort of abstracting agencies, Herner and
company (...) conducted a survey (...) The
purpose was to determine if and how... (J 412).

(65)


```
                    SO -
               /            \
              C              G
              |          +Time sequence
             S1+             |
              |             S2+
             / \           /    \
            /   \         C      G
           /_____\        |      |
                         S1'+   S2'- ?
                          |      |
                         / \    / \
                        /___\  /___\
        such cards can...   The next step  to select...
                            is to search..
```
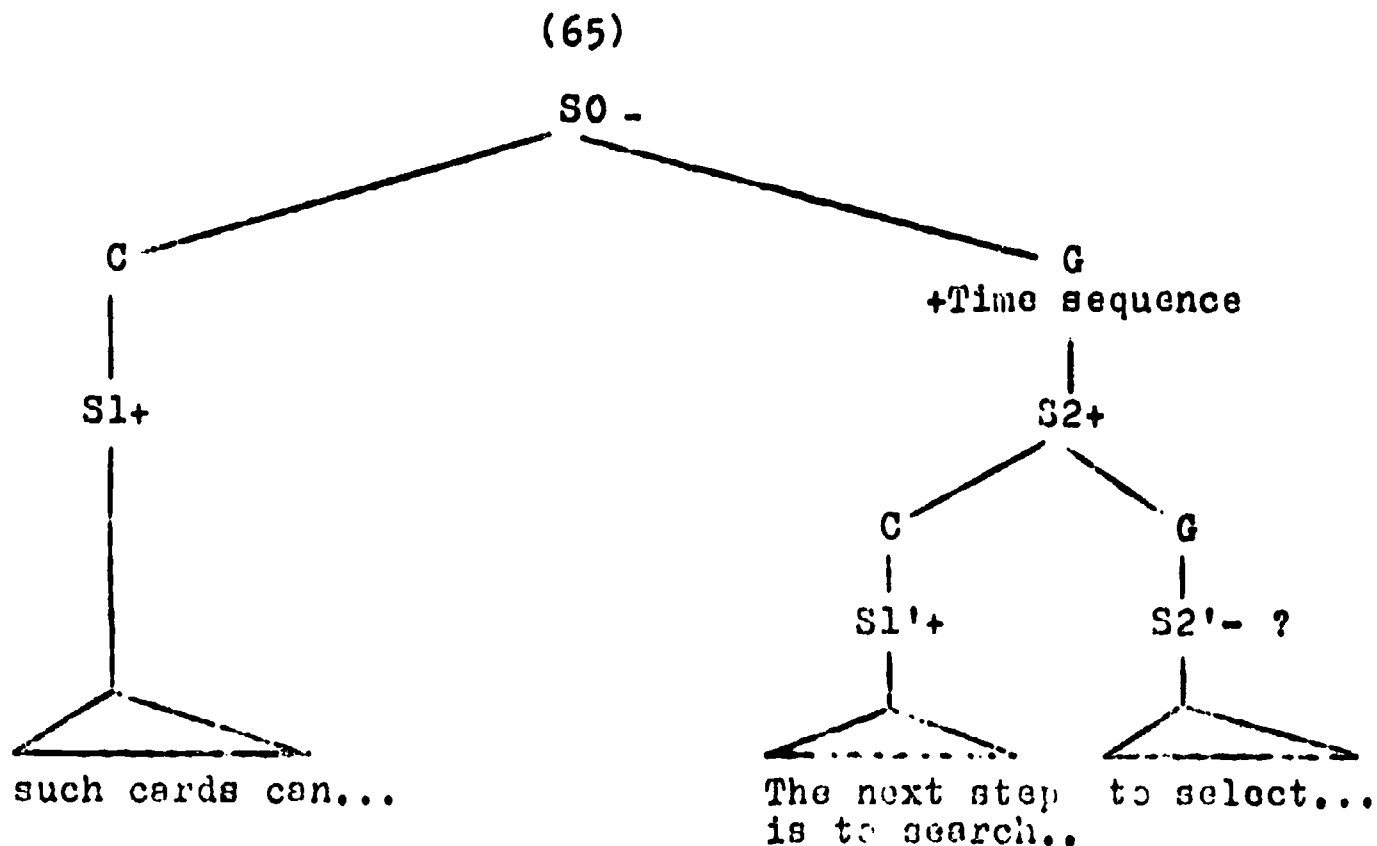
Example : Such cards can easily be turned into a
magnetic tape. The next step is to search
the tapes for fractional structures and so
to select chemical types of a predetermined
character from the stock (J 273).
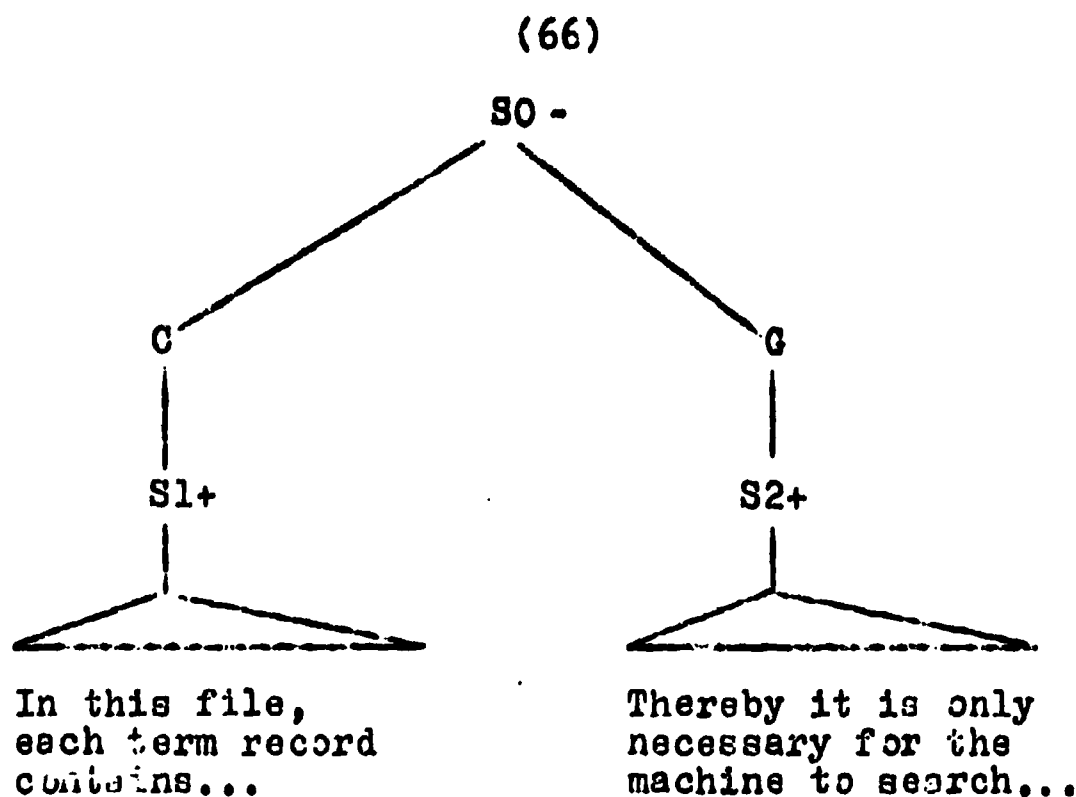
(66)

```
                            SO -
                     ╱              ╲
                    C                G
                    │                │
                   S1+              S2+
```

In this file,   Thereby it is only
each term record   necessary for the
contains...   machine to search...

Example : In this file each term record contains
all of the document codes for all of the
documents in which the term appears. Thereby
it is only necessary for the machine to
search the pertinent terms in the machine's
memory (J 9).

(67)

```
                            SO-
                     ╱              ╲
                    C                G
                    │                │
                   S1+              S2-
```

... the classification   thereby making it possible
system is made from...   to search...
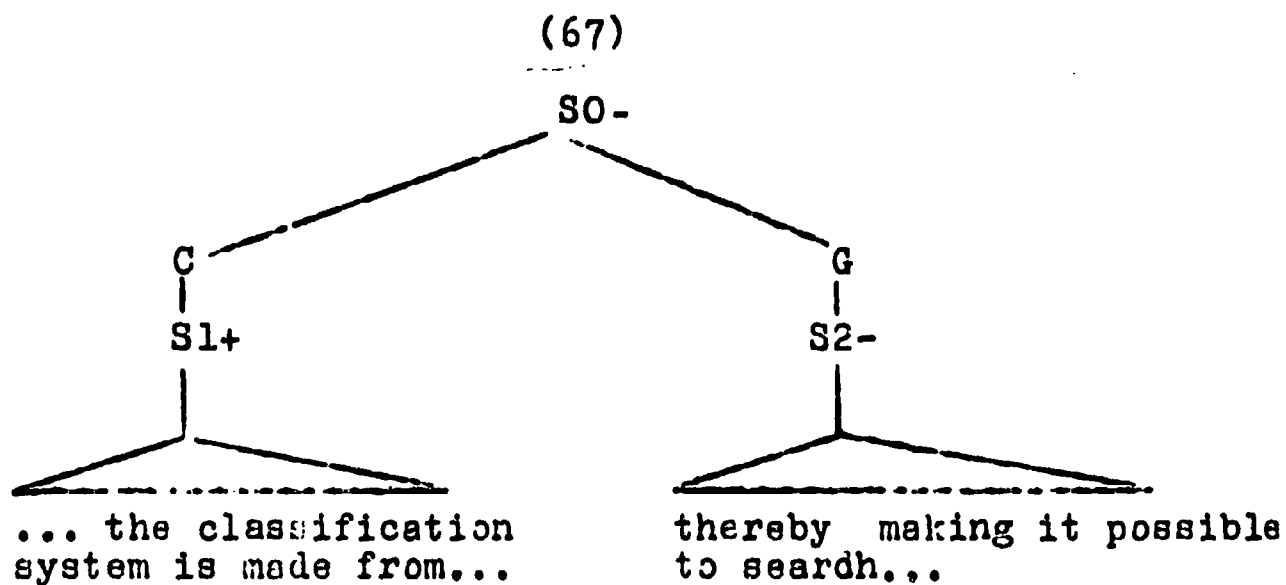
Example : In this system, the classification system
is made from the disclosures in the patent
literautre, thereby making it possible to
search for all the information in the
file of interest to the searcher (J 19).

(68)


```
                        SO-
              /                    \
             C                      G
          +Reason                   |
             |                     S2+
            S1-                 /        \
             |                 C          G
             |                 |     +Time sequence
             |                S1'+         |
             |                 |          S2'-
    /‾‾‾‾‾‾‾‾\       /‾‾‾‾‾‾‾\     /‾‾‾‾‾‾‾‾\
    simply because   ...search...   prior to embarking..
    library research  ...... b.
    costs less...     ......
```
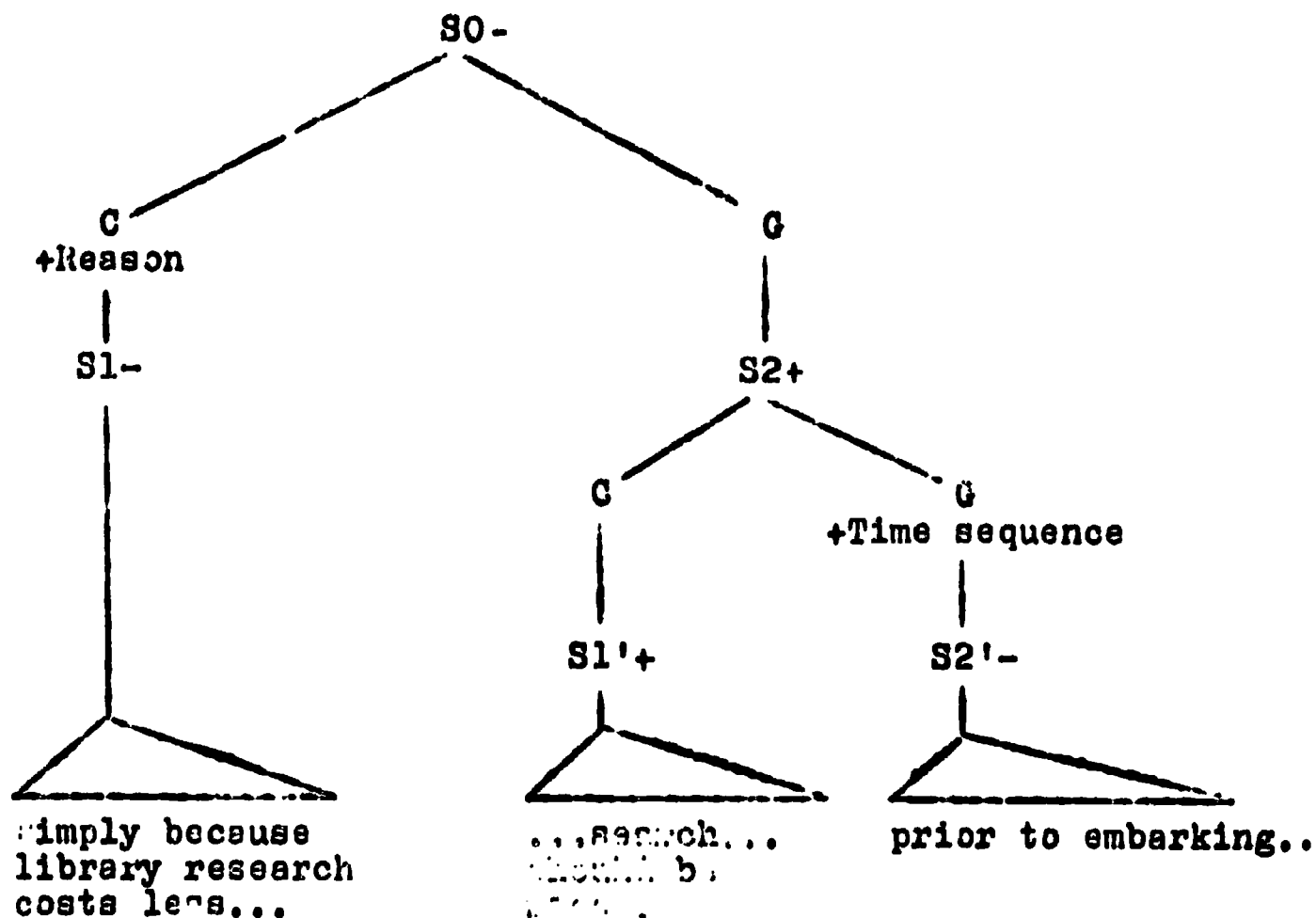
Example : Prior to embarking on a laboratory project,
a thorough search of available literature
should be made, simply because library
research costs less than laboratory research
(J 889, sentence 1).

(69)

Alternative analysis of S2 in (68)


```
                    S2++
          /          |          \
         V           C           G
         |           |           |
         |          S1'-         S2'-
         |       /‾‾‾‾‾\      /‾‾‾‾‾‾‾\
    should      search be made   prior to embarking...
    (cp. necessary)
```

(70)

$$SO-$$

```
        C                      G
        |                      |
       S1-                    S2+
```

With the volume of   this is becoming
literature doubling...   more and more difficult..
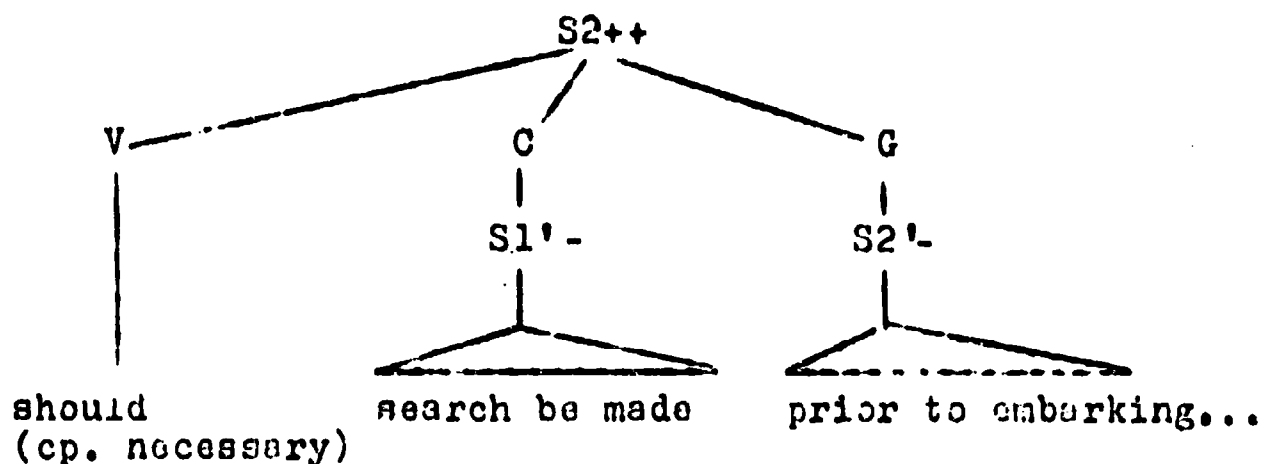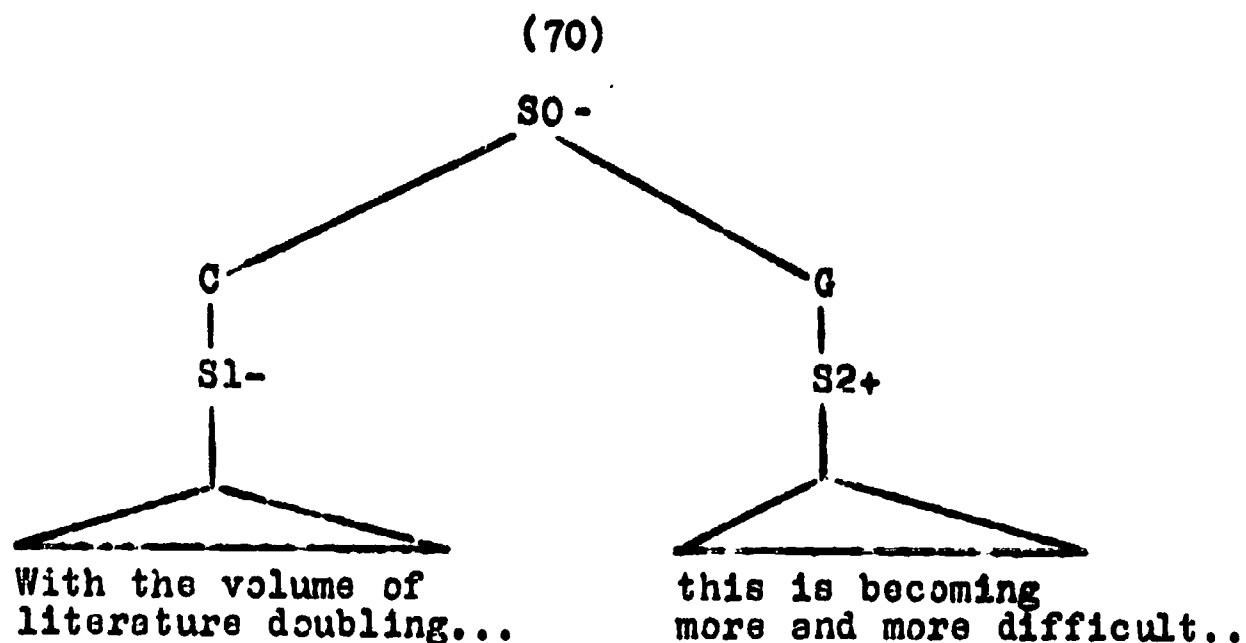
Example : (However,) WITH the volume of literature
doubling every ten to fifteen years, this
is becoming more and more difficult
(J 889, sentence 2).

The last trees are intended to suggest a possible
relevance of my analyses to discourse structure. Thus,
bringing (68) and (70) together, the first two sentences
of J 888 would be represented as follows, ignoring the
problem of how to represent 'however' :

(71)

$$S$$

$$SO^1- \qquad SO^2-$$

```
   C            G            C          G
   |            |            |          |
  S1-          S2+          S1-        S2+
            C       G
            |       |
          S1'+    S2'-
```

because   search should   prior to   With litera   However,S2 of
library   be made   embarking... ture doubl-   SO^1is becoming
research costs   ing....   more and more
less   difficult.
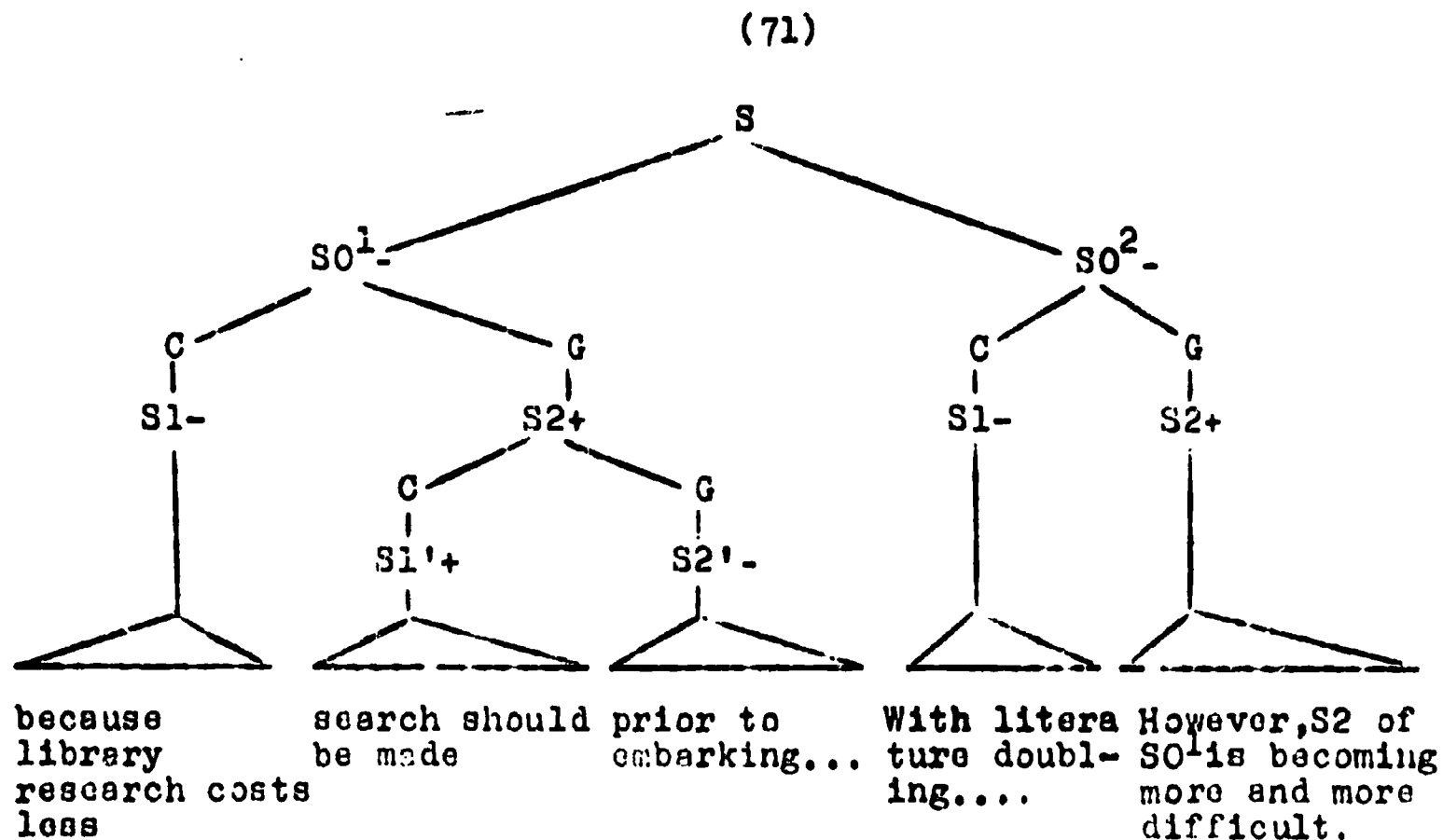
A similar application to discourse structure is possible in the first two sentences of J 390; in the second sentence which is diagrammed below, <u>this</u> refers to the kernelization of texts mentioned in the first sentence and can be analyzed as C :

(72)

```
                        SO+
           _____/ |  _____
          /            C               G
         V             |               |
         |             NP              S2-
         |             |               |
      provides       this      ____/___|_____
                              for easier storage
                              and retrieval
```
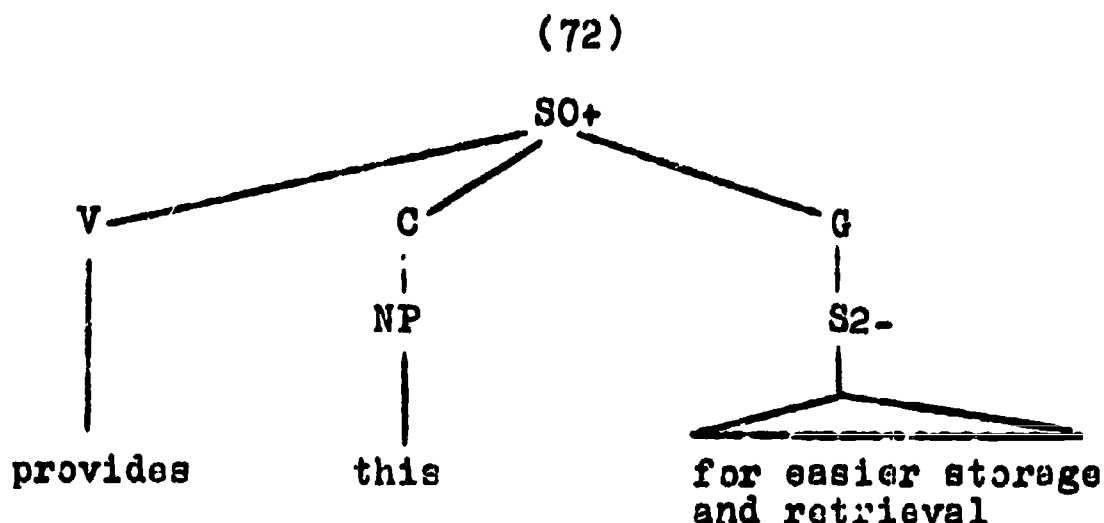
<u>Example</u> : (Concerned with the idea that) a text (...) may bo reduced to a sequence of kernel sentences which convey the same information as the original piece of work. This <u>provides for</u> easier storage and retrieval (J 390).

The above example suggests two remarks. First of all, suppose we had '<u>facilitates</u>' instead of '<u>provides for easier...</u>.' This presents a difficulty for case theory as it currently stands : '<u>storage and retrieval</u>' would presumably have to bo analyzed as the Object of '<u>facilitates</u>'. The only way to correct this deficiency while preserving the general principles of case theory is to resort to meaning postulates : for instance, PROVIDE (C, G = <u>easier S</u>) = FACILITATE (C, O = S). Similar postulates would be needed for other causative verbs such as '<u>simplify</u>' (= MAKE + Goal = <u>simpler S or NP</u>) as in the following example : "Advances in linguistics could greatly <u>simplify</u> the unsolved and growing problem of formulating and communicating new knowledge" (J 1027). In fact, this property of causatives seems to bo so general and causatives are such an important class of verbs that some kind of general rule similar to a redundancy rule seems to be necessary. This rule would state

that Causative + Object implies CAUSE + (the corresponding non-causative + Object (or, if necessary, some other case)) as a Goal.

Second, in the example of (72) ("This provides for easier storage and retrieval"), 'This' stands for a sentence. This suggests that in many of the analyses so far presented (particularly those marked SO + because their surface form has a predicate such as "provides") what has been written as S should in fact be written as NP = S, that is NP rewritten as S. In such cases then, the S I have used so far must be regarded as a notational simplification.

To conclude, I would say this . The analyses proposed in this and the preceding section are obviously different from those so far proposed within case theory. Except in sentences of the type "The purpose of— is to—" (39), I see no reasons for extending the system I propose to the type of 'simple' sentences usually analyzed by Fillmore and his followers, or for giving up the type of representations they propose for 'simple' sentences. I shall therefore assume the validity of these representations in what follows. What I do feel, however, is that the type of representation I propose is justifiable by the generalizations it captures. Moreover, the 'simple' sentences find their place in my system as exponents of S1 and S2, in which Fillmore's case structures can in turn find their place. Evidence for such a conception is the following sentence (73), which was analyzed above in (61). As far as I can see, Fillmore would analyze the by- and the with- phrases in this sentence as Instrument. But this would result in a violation of his principle that "each case relationship occurs only once in a simple sentence" (Fillmore 1968a, 21), since another Instrument ("by the computer"), which I cannot see how to relate to the by- and with-phrases, is associated with the verb of the main sentence ("are selected"):

(73)

By noting the number of matching terms between this
extended list of request terms and the terms used in
indexing a document, and **with** due regard for their
degree of association, documents are selected **by the
computer** and arranged in the order of their relevance
to the request (J 942).

Under the analysis I propose, this sentence is not
a violation of Fillmore's principle.

3.5. SOME COMPLEX INSTRUMENTS.

3.5.1. "Machine" and "system" as Instruments.

Case theory as it currently stands recognizes the following types of Instruments :

(1) 'Simple' Instruments : for instance, 'the key' in "The key opened the door" and in "John opened the door with the key" (Fillmore 1968, 25).

(ii) Corresponding constructions with the verb 'use', as in "John used the key to open the door" (loc. cit.). Fillmore, however, has never published an analysis of such sentences. He has merely suggested in lectures that the grammatical object of an active sentence with 'use' is to be analyzed as an Instrument (Fillmore 1970 b). Langendoen. who has adopted some of Fillmore's proposals, suggests to analyze the verb 'use' as "a three-place predicate, whose arguments are an agent, an instrument, and an infinitival clause expressing what may be called result". This clause in turn is analyzed as containing a predicate (e.g. 'open'), an agent and another case (e.g. the object 'door'). Langendoen proposes to capture the relatedness between the sentences with 'use' and corresponding sentences of the type given in (1) by assigning them the same structure, the only difference being that sentences such as "John opened the door with the key" have an abstract instrumental predicate without lexical realization (Langendoen 1969, 106-7).

(iii) Instruments such as the subject of 'proved' in "The fact that he had blood on his hands proved that he was guilty" (Stockwell et al. 1968, 10, cp. Fillmore 1970 b.

(iv) Compound instruments, which are described as "containing a <u>possessed noun</u> as the Instrument (<u>the car's fender</u>)", and as allowing two possibilities of subject selection: compare "The car broke the window <u>with its fender</u>" and "<u>The car's fender</u> broke the window". "A similar explanation is suggested for" : "Your speech impressed us with its brevity" and "The brevity of your speech impressed us" (Fillmore 1968, 23).

In this chapter, I wish to discuss types of Instrumental constructions which to my knowledge have never been examined within case theory, as well as some conceptual dependencies between Instruments, which are in certain respects similar to, and in other respects different from Fillmore's example with "<u>the car's fender</u>". The need for such a discussion in this study will become apparent as soon as one remembers Webster's definition of '<u>documentation</u>', which is repeated in (1) (a), and a possible paraphrase of it given in (1) (b), in which I have capitalized the Instrumental predicates and underlined what would presumably be analyzed as Instrument by Fillmore :

(1)

(a) the assembling, coding and disseminating of recorded knowledge comprehensively treated as an integral procedure utilizing semantics, psychological and mechanical aids, and techniques of reproduction including microcopy for giving documentary information maximum accessibility and usability.

(b) a <u>procedure which</u> UTILIZES <u>semantics, psychological and mechanical aids,</u> and techniques USED for reproduction, and <u>which</u> is USED for the assembling, coding and disseminating of recorded knowledge, and for giving <u>documentary information</u> maximum accessibility and USABILITY.

In this section, I shall be concerned with the status of 'machine', 'system', and equivalent words as Instruments. I shall make use of five criteria for establishing che Instrumental status of these words. The first two criteria have been used in combination by Fillmore and others; I believe, however, that the other three are equally useful.

(a) 'Machine', 'system', and equivalent words function as subject of various verbs.

(b) In prepositional phrases, they are introduced by prepositions that are typically associated with Instruments in English.

(c) They function as grammatical object of verbs like 'use' in active sentences.

(d) There are morphological formations with -er and -or which are ambiguous between Agent and Instrument in the same way as are grammatical subjects. In the Instrumental reading, these suffixes are equivalent to 'machine' or 'system' : for instance, "technical translator" (J 438, Agent), versus "automatic translator" (J 724, Instrument). Compare the following advertisement : "Don't BE a dishwasher, BUY one".

(e) Lastly, there are some nominal groups in which the Instrumental analysis seems to be relevant to word order.

(a) to (e) must be regarded as a 'battery' of tests, in which each separate test or criterion is necessary next to the others. Evidence for the relatedness of the five constructions can be found in the existence of texts like the following (2 a and b), in which two or more such constructions co-occur and have a related meaning. The fact that 'machine' and 'system' can be conjoined (as in 2 e) { evidence for their relatedness. In (2) (d), 'method' is used with the preposition 'by' in an active sentence, and as object of 'employ'.

(2)

(a) A review is given of the progress in research and de-
velopment on the application of high-speed digital
computers to language translation (e.g., Russian
to English). An attempt is made to determine (1)
how machine translation can be accomplished, (2)
what the principal achievements are thus far, and
(3) what more is required. Machine translation
(MT) is defined as the art of translation from one
natural language into another, by means of an auto-
matic computer (J 276).

(b) On the recognition of information with a digital
computer.
The author analyzes the problem of data recognition
by computer, in particular for "table look-up" pro-
cedures, taking into account the questions of per-
missible error by the computer and the permissible
errors in data which the machine may be required
to recognize, in the nature of NW YRK as recog-
nizably New York. The author concludes that a
completely comprehensive recognition system,
allowing for all possible errors, will probably prove
uneconomical. He suggests that the root of the
problem is in the necessity for the computer to
examine each character individually and serially,
as opposed to the human capacity to scan a whole
word or phrase for recognizable patterns. He sug-
gests further that ultime solution may lie along the
lines of an analog based on the harmonic patterns
of the word or phrase (J 340).

'c) Machines and systems for the modern library.
The major portion of this paper is devoted to an
examination of the fundamentals involved in the
various existing machines, and systems as applied
to modern libraries, and to descriptions of the ma-
chines and systems and "their areas of applicability".
Among the systems discussed are Batten on Peek-a-
Boo cards, the Coordinate Indexing system, tele-
graphic code, and the system used for indexing the
field of steroid chemistry. (J 1065).

(d) Methods by which research workers find information
(...) what methods scientists employ in obtaining
information... (J 298).

In the examples listed below, I have found no decisive
evidence for or against an analysis of 'machine' or 'system'
as Cause, the case label proposed by Langendoen for "instrument-
like expressions" in sentences having no "understood agent"
(Langendoen 1970, 74). I have found no example in which it
was possible with any certainty either to rule out or to
establish the presence of an implicit Agent. I shall therefore
not examine the possibility of a 'Causal' analysis of 'machine'
and 'system' in this section.

Let us first consider a few attested examples of
'machine' and 'system' (and occasionally some equivalent words)
functioning as subject.In the examples of (3) 'machine' is the
superficial subject and in the examples of (4) it is the
subject of a clause in which Equi-NP deletion can be thought
to have operated. The word 'machine' and the verb of which
it is the subject are underlined in the examples. In (4).
I have avoided quoting examples with 'for' + infinitive
because they may be interpreted in two ways and therefore
provide no clue for the status of 'machine' as subject. Thus,
'machine for scanning...' is relatable to 'machine for people
to scan... (with)', as well as to 'machine which scans...'.
Also compare what I believe to be an indication of the
Instrumental status of 'machine' in (4d) ("the machine is
able to do...") with the following example which I believe
to be non-critical : "This punched-card system enables
information to be retrieved..." (J 440). Similar examples
with 'system' as subject are given in (5) and (6).

(3)

(a) No __machine__ now commercially available __can do the job__ [__of indexing and searching__] efficiently (J 444).

(b) This paper attempts to develop formal distinctions between a __machine__ that passively __classifies__ its inputs into predetermined categories and one that __develops__ its own classificatory scheme (J 483).

(c) A __machine__ that does research (J 500).

(d) The __machine can search__ for any combination of a number of coded characteristics or even for their absence (J 504).

(e) This __machine__, called the universal card scanner, __scans__ cards ... (J 609).

(f) The __machine__ for __scanning__ these encoded cards __can search__ up to 700 cards per minute (J 854).

(g) ... the fears that __machines will mechanize__ the human mind... (J 866).

(h) __Machines__ alone __can__ never __solve__ the fundamental problems of information storage and retrieval (J 344).

(i) In such a system, __machines__ are powerful tools which, under the guidance of human judgement, __can do routine operations__ that free the creative talents of man (J 436).

(j) ...__machines__ which __can provide__ essays on any given subject upon request (J 683).

(k) ... each __machine performing__ efficiently one of the stages in the operation (J 684).

(l) Thereby it is only necessary for the __machine__ to __search__ the pertinent terms in the machine's memory (J 9).

(m) A __computer__ which finds the subject category...(J 639).

(4)

(a) ... the __ability__ of a __machine__ to adequately __determine__ information relevance depends strongly on its ability to __recognize__ and manipulate the syntactic structures of the text (J 176).

(b) ... the permissible error in data which the __machine__ may be required to __recognize__... (J 340).

(c) An electronic <u>machine</u> that <u>answers</u> the requirements of information retrieval by <u>scanning</u> of punched-cards... (J 609).

(d) However, this <u>machine</u> is <u>able</u> to <u>do</u> more than the <u>scanning</u> and <u>selecting operations</u> described (J 770).

(e) A description of an automatic microfilm <u>searching</u> <u>machine</u> designed to <u>increase</u> searching speed, to <u>edit out</u> extraneous matter, and to <u>provide</u> copies of the desired information (J 1109).

(f) ... there is a need for both the high-speed input <u>machines</u> to <u>handle</u> the bulk of the text transcriptions.. (J 679).

(g) data which the <u>machine</u> may be required to <u>recognize</u> (J 340).

(h) the necessity for the <u>computer</u> to <u>examine</u>... (J 340).

(i) the <u>machine</u> as an <u>instrument</u> for <u>retrieving</u> information (J 943).

(5)

(a) This <u>system handles</u> large numbers of chemical compounds and polymers... (J 275).

(b) ... there is no one <u>system</u> that will <u>handle</u> all knowledge (J 296).

(c) Describes a <u>system</u> which has these features : <u>potential for</u> generic, specific, or combination <u>searching</u> (...) <u>searches</u> compound, process, mixtures and biological data... (J 318).

(d) ... the <u>system</u> would <u>search</u> a million documents (J 689).

(e) <u>system</u> ... which <u>discloses</u> any subject matter (J 704).

(f) <u>system</u> ... <u>working</u> well ... (J 718).

(g) a <u>system</u> of coding by symbolic logic which will not just <u>retrieve</u> information but also <u>create</u> new ideas (J 846).

(h) ... the uniterm <u>system gave</u> slightly better results... (J 872).

(i) what such a <u>system</u> should <u>provide</u> (J 938).

(j) The <u>system cut</u> indexing time, <u>increased</u> retrieval speed, and <u>allowed</u> greater depth of indexing than the subject heading list which had been used previously (J 1119).

(k) Of the two output **Systems** designed, one **stores...** and **displays** it... the second **reproduces**...(J 37).

(l) no large scale information retrieval **system** can be a success unless it **can handle searching** for complex relationships (J 20).

(m) **system** which **can find** documents (...) and **present** these documents (J 942).

(6)

(a) this structure enables the **system** to : directly **interrogate**... (J 494).

(b) This **system** is designed to **provide multiple access to** ... (J 570).

(c) This **system** is designed to **minimize** (...) and **maximize**.. (J 758).

(d) code **system** developed to **represent** chemical structures (J 707).

(e) **system** devised (...) to **accomodate** a collection ... (J 848).

(f) **system** (...) established to **retrieve** items...

Examples of uses of '**machine**' and '**system**' and occasionally equivalent words with 'instrumental' prepositions are given below in (7) and (8), and examples of the same words associated with the verb '**use**' and equivalent words are given in (9) and (10). Examples with the preposition '**by**' have been included in (7) and (8) when the absence of the article and the possibility of a parallel active construction with '**by**' or '**with**' suggests that we have to do with an instrumental '**by**' rather than with the by-phrase of the passive.

(7)

(a) Machine translation is defined as the art of **translation** from one natural language into another **by means of** an automatic **computer** (J 276).

(b) the information can be **retrieved by** machine (J 321).

(c) Pattern **recognition by** machine (J 883).

(d) **simulating** thought processes **by** machine (J 953).

(e) <u>recognition</u> of information <u>with</u> a digital <u>computer</u> (J 340).

(f) literature <u>searching</u> <u>by</u> <u>machines</u> (J 614).

(g) <u>education</u> <u>by</u> <u>machine</u> (J 683).

(h) The coded information is <u>searched</u> <u>by</u> <u>machine</u> (J 22).

(i) <u>Finding</u> chemical records <u>by</u> digital <u>computer</u> (J 819).

(j) literature searching <u>by</u> <u>machines</u> (J 614).

<center>(8)</center>

(a) <u>By means of</u> this system, 't is also possible to <u>search</u> ... (J 273).

(b) Successful <u>searches</u> <u>by</u> this mechanized <u>system</u>... (J 517).

(c) can things be <u>accomplished</u> through using machine-sorted cards which cannot be accomplished <u>through</u> <u>other methods</u> (J 95).

(d) ... to improve <u>searching</u> <u>with</u> the peek-a-boo filing and retrieving <u>system</u> (J 1017).

(e) the desired information is very readily <u>located</u> <u>by means of</u> a simple and practical <u>system</u> (J 1092).

(f) The technique was devised by Stern to improve <u>searching</u> <u>with</u> the peek-a-boo filing and retrieving <u>system</u> used in the reference service (J 1017).

(g) Three thousand structural formulas for encoding (...) were divided into 3 groups of 1000 each. Each group of 1000 compounds was <u>encoded by each of the 2</u> <u>systems by</u> one each of 3 practising chemists, 3 students working toward the doctorate in chemistry, and 3 undergraduate students majoring in chemistry (J 797).

(g') x Each group of 1000 compounds was encoded by each system and by practising chemists.

<center>(9)</center>

(a) the <u>machine</u> should be simple to demonstrate and <u>use</u> (J 51).

(b) the indexed information is retrieved by <u>using</u> an IBM-101 electronic statistical <u>machine</u> (J 275).

(c) <u>Application</u> of high-speed <u>computers</u> to information retrieval (J 18).

(d) <u>Application</u> of <u>machines</u> in the searching process (J 99).

(e) by 1961 these <u>machines</u> will <u>serve</u> as translators (J 624).

(10)

(a) The <u>application</u> of a punched-card <u>system</u> for <u>indexing</u> rocks and minerals (J 830).

(b) The <u>application</u> of (...) a new classification <u>system</u> to the <u>preparation</u> of an analytical index (J 1111).

(c) <u>system</u> now in <u>use</u> (J 63).

(d) <u>use</u> of the Coden <u>system</u> by the individual research scientist (J 91).

(e) how such a <u>system</u> would <u>operate</u> (J 473).

(f) the way in which the <u>system is operated</u> by the Westinghouse Electric Corporation (J 35).

(g) the peek-a-boo filing and retrieving <u>system used</u> in the reference service (J 1017).

One particularly interesting example is (8g). The critical sentence in this passage shows a clear contrast between instrumental and agentive by-phrases, one following the other, and the sentence corroborates Fillmore's postulate that "only noun phrases representing the same case may be conjoined" (Fillmore 1968a 22).

The facts that are available to me about nominal compounds in <u>-er</u> and <u>-or</u> suggest that, contrary to what a 'lexicalist' may say, they require an abstract treatment such as that first outlined in Lakoff (1968 a), <u>and</u> an analysis in terms of case theory to capture the obvious contrast between the Agentive and the Instrumental meaning of such compounds. My observations here are in agreement with Lakoff's proposal to posit, say, for '<u>mixer</u>' an underlying structure of the type "a device with which one mixes things and to have an Instrumental nominaliza-tion rule deriving nominal compounds in <u>-er</u> or <u>-or</u> for the verbs that allow such a derivation (G. Lakoff 1968a, IV-3). The only counterevidence to the kind of analysis proposed by Lakoff and since then extended by Bach 1968 to all nouns is the existence of expressions like "<u>a human indexer</u>"(J 632 ;

with NO corresponding attested example of "indexer" with an Instrumental meaning). I am unable to explain the occurrence of 'human' here which obviously duplicates 'one' if we posit a deep structure of the type "one who indexes" for 'indexer'.

For the rest, I have found an obvious correlation between the -er and -or suffixes and words like 'machine', 'device' and 'system'. The best-known example is of course "the computer", often referred to as 'a(n electronic) data processing system' or as 'the machine'. Additional evidence for this correlation can be found in various passages like the following (11), in which "system" or some other word such as "element" is relatable to, or refers forward or backward to a word in -er or -or. As is shown in (11 h), these compounds can be used as direct object of the verb 'use', like 'system' :

(11)

(a) Details are given of the Copyflex high-speed printer, and the way in which the system is operated...
(J 35).

(b) Advantages of the collator over manual systems (J 1083).

(c) the replacement of alphabetical and numerical coding systems by word indicators (J 235).

(d) the insertion of a machine as storage element and/or selector between contributor and user (J 215).

(e) A description of the Index Searcher (...) which includes a list of 8 design features of the system (J 517).

(f) This machine, called the 'Universal Card Scanner' (...) scans cards... (J 609).

(g) Fact Compiler : a system for the extraction, storage and retrieval of information (J 494).

(h) using punched-card sorters and collaters (J 963).

Various other points would be worth investigating in connection with _-er_, _-or_, and other suffixes (e.g. "Find_ex_": J 830). The most important point for the purposes of this discussion is the contrast between Instrument and Agent. Compare, for instance, "automatic/mechanical translat_or_" (J 724, 829, 953) and "technical translat_or_". (J 438)(cp. pr_inter_). Occasionally, there exists an Agentive compound next to morphologically different or unmarked Instrumentals, as in "_indexer/or_ (Agent) versus "_descriptor_" and "_index_" (cp. _store_: _memory_; _reservoir_, J 851). Also note that some Instrumental compounds seem to have no Agentive counterpart : for instance "(the golden) _retriever_" occurs only once (J 296), and with an Instrumental value (it is not the dog of our desk dictionaries !). Another morphologically unmarked Instrumental is '_punch_' as in "IBM 026 Card Punch" (J 273) and in "Keypunch" (J 728), but "punch _machine_" (J 728) is also attested.

In most of the Instrumental compounds, the _-er/or_ ending corresponds to words like _system_, _machine_, _tool_, etc. : compare "Scanner" (J 609) and "scanning _tools_" (J 629). In some cases, however, the ending corresponds to a more specific word : a "descript_or_", for instance, can be defined as "_a key word or phrase_ used in data processing to identify an item (as a subject or document)" (Webster's Third Addenda Section). It seems then that the only constant corresponding to the suffix is just the Instrumental role of this ending or of the equivalent word. But now, consider the following attested example : "Research is being done on the development of special small-size _machine-translators_" (J 624). The analysis I would propose is the following. Since the _-or_ ending is equatable to '_system_', the above nominal group can be paraphrased as '_machine translating system_'. This is indeed a type of construction which is

well attested in the Janaske corpus, namely machine + Ving (or a cognate nominal) + system : for instance, "machine indexing system" (J 401), and other related examples which I shall quote presently. Note, moreover, that while machine + Ving, and Ving + machine are possible and indeed frequent (machine indexing/searching and searching/translating machine, etc.), only Ving + system is possible; system + Ving, as far as I can see, does not exist, and is not attested in the Janaske corpus. This suggests the existence of a linguistically marked hierarchy among Instruments. I shall discuss this topic below (cp. "seven system models", J 681, and "On retrieval system theory", J 1044).

One last argument in favour of an Instrumental analysis for 'machine', 'system' and equivalent terms is the existence of nominal groups, the surface form of which is given in the following table (12). The first position cann be filled by nouns which can be marked [Human] : for instance , x man (information) retrieval, or x indexer (information) retrieval. On the other hand, for a reason which I am unable to explain, 'human' IS possible in the position usually occupied by Instruments. So are the adjective 'manual' and the corresponding noun 'hand' in the same position. Particularly the last example given in (12) suggests that 'manual' and 'hand' need to be treated as Instruments, like 'machine'. An Instrumental analysis for 'human' may also be considered for examples like the following :"print-out for human searching by eye " (J 256). Such a treatment raises various problems, some of which will be discussed below (cp. Fillmore's treatment of an implicit 'body' as both Instrument and object in "John arose": Fillmore 1969b, 136, fn. 8).

(12)

Instrument 1	+ Object	Ving or cognate Nom.	+ "system"	
machine	information	storage and retrieval		J 646
machine	reference	retrieval		J 580
machine		indexing		J 450
machine		retrieval		J 461
machine		translation	techniques	J 766
all-computer	document-	retrieval	system	J 942
machine	literature	searching	systems	J 775
electronic				
computer	literature	search		J 756
x man		sorting		
human		searching		J 256
manual		searching	tools	J 413
manual		sorting		J 472
hand-		sorting		J 529
manual needle		sorting		J 790
(from)a manual				
to a machine		indexing	system	J 401

3.5.2. Some problematic examples.

I now wish to examine various situations in which the analysis of 'system','machine', and various other words as Instruments of certain predicates is problematic for a number of reasons. More broadly, I shall discuss some problems arising from the central, axiomatic role which Fillmore assigns in case theory to the [± Animate] distinction in selecting certain nouns or noun-phrases as either Agents or Instruments.

I shall first return . for a while to Chomsky 1965. According to the Aspects theory, "John frightened sincerity" is described as "a deviant sentence, formed by relaxing the restriction of frighten to Animate Direct-Objects. Nevertheless, Chomsky feels, there are frames in which this restriction can be violated with no consequent unnaturalness". As examples, he gives various sentences with negations : for instance, "it is nonsense to speak of (there is no such activity as) frightening sincerity", "one can(not) frighten sincerity".Chomsky does not state the (syntactic and semantic) rules by which the "grammar" would indicate that, while 'John frightened sincerity' is deviant, sentences like the above "are not" (Chomsky 1965, 157). Note that in this particular discussion, Chomsky does not allow for intuitions different from his own : for instance, he does not imagine that some people could find it natural to personify "sincerity" in "John frightened sincerity". In fact, elsewhere he suggests to account for personification in terms of violations of selectional rules, which seems to suggest that to him such facts as metaphor are to be handled by the semantic component but are to be described as deviant by the syntax (op. cit., 149). Fillmore, as I have pointed out in 3.1., does grant personification a place in his case grammar, but unfortunately

he does not work out any of the details. The same is true of
Lakoff's more revolutionary suggestion to set up a theory
of "relative grammaticality" which would generate representa-
tions of sentences together with the (representations of)
the beliefs that make these sentences acceptable - even
when they can be found unacceptable under different beliefs
or in different contexts (G. Lakoff 1969b and 1971).

Here, I shall adopt George Lakoff's attitude rather than
Chomsky's, and I shall examine the question of what extensions
of case theory are needed to introduce relative grammaticality
into it. Making provisions for different, and occasionally
conflicting, beliefs in the theory is necessary in any case,
to account for a host of facts which would otherwise have to
be ruled out or to await explanation in some suitable component
of the theory (for instance, sentences like "My cat enjoys
tormenting me"), AND to account for systematic correlations
between such facts (e.g. gender) and matters of belief
(Lakoff 1969b, 104-109). Furthermore, if one adopts Lakoff's
attitude, the problem of sentences with negations, as are
discussed by Chomsky, can be reduced to the problem of sentences
without such negations. In fact, it seems more plausible to
do it that way than the way Chomsky proposed to do it : one
may wonder what is being denied by the negation in "sincerity
is not the sort of thing that can be frightened" if "frighten
+ sincerity" is labelled a priori as deviant (Chomsky, loc. cit.)
On the basis of what has just been said, I shall ignore the
problem of negation in what follows, in order to concentrate
on points relevant to case theory.

The examples from the Janaske corpus which I now wish
to discuss have been selected as problematic because they
fail to pass two tests which could have been used in the
preceding section to establish my Instrumental analyses. As
will be seen, my tests cast doubt on the assumption, commonly
made in the literature of TG linguistics, that sentences of
the type "x uses y to do something" and "x does something with
y" are to be given the same deep structure (G. Lakoff 1968b,
Fillmore 1968a).

(1) The first test is based on the verb 'use'. Consider the
following sentences :

(13)

   (a) As a final comment the author adds that one should
      remember that the machines developed to date
      are only able to retrieve what has been put into
      them, nothing more and nothing less.

   (b) The machine is neither able to "think" about the
      information put into it nor able to "forget"
      obsolete information.

I cannot help feeling that, while (14) (a) below is
relatable to (13) (a) and means that the machine does the
retrieving, (14) (b) and (c) mean that 'the man' does
the thinking and the forgetting, and are NOT relatable
to (13) (b) :

(14)

This man uses machines :
(a) to retrieve the information put into them.
(b) to think about the information put into them.
(c) to forget obsolete information put into them.

Also note that (15) (a) is acceptable whereas (15) (b) and
(c) to me are dubious (the with-test is to be found in
Chomsky 1969) except perhaps in some queer science fiction
context :

(15)

   (a) He uses machines to retrieve information <u>with</u>.
   (b) ? He uses machines to think about information <u>with</u>.
   (c) ? He uses machines to forget information <u>with</u>.

From this, I would conclude that '<u>think</u>' and '<u>forget</u>'
cannot have '<u>machine</u>' or the like as Instrument. (No
more than Chomsky am I able to provide a justified account
for the <u>with</u> in (15 a) ; sentences like "x He uses machines
to retrieve information <u>with machines</u>" are not acceptable
whereas "He uses machines to retrieve information <u>with them</u>"
may be; this suggests that an Instrument could be posited
in both the '<u>use</u>' and   the <u>to</u>-clauses).

(11) The second test makes use of the word '<u>application</u>'. I
feel that while (16) (a) sounds quite natural, I am
unable to understand (16) (b) and (c) :

(16)

   (a) his application of machines to the betrieval of
       information
   (b) x his application of machines to thinking about
       information
   (c) x his application of machines to forgetting
       information.

If I am correct in assuming that verbs like '<u>think</u>'
cannot have '<u>machine</u>' and the like as Instruments, then the
selection of the subject of sentences like (13)(b) needs to be
made sensitive to differences in people's beliefs. Thus
according to some people, only human beings can think, but
according to others machines are also able to think. Within
current case theory, such differences can only be captured
by <u>both allowing and disallowing</u> NP's with '<u>machine</u>' and the
like to function as AGENT of sentences with predicates like
'<u>think</u>'. I am not familiar with any attempt within case theory
to set up feature-changing rules that would, for instance,
change a feature [- Animate] into [+ Animate]. What I would

therefore propose is a form of lexical entry allowing for mutually incompatible, or contradictory, feature-specifications: thus, a noun like 'machine' would be marked both [- Animate] and [+ Animate] so that its selection as Agent with verbs like 'think' would be both allowed and disallowed, depending on what the speaker wishes to say, and on his beliefs (about 'machines', about 'thinking', etc.).

Because I believe that the underlined verbs fail to pass the two tests just mentioned, I feel that 'machine' furctions as Agent in the following sentences given in (17). Note that only the last two passages contain no negation. Also note that some passages are clearly beyond the powers of current theories(for instance, the sentence with intelligence in 17 a); this, however, does not affect my argument.

(17)

(a) The machines do not think but they can now meet nearly all the criteria of intelligence except pattern recognition (J 895).
(b) The myth of thinking machines... premature announcement of what someone expects either he or a machine will do but has not done (J 964).
(c) The machine does not possess a will, and its so-called "conclusions" are only the logical consequences of its input (J 857).
(d) A progress report on machines to learn to translate languages and retrieve information (J 922).
(e) As machines learn they may develop unforeseen strategies at rates that baffle their programmers [= title]. The writer develops the thesis that machines can and do transcend some of the limitations of their designers (J 1096).

Also note that my proposed analysis of 'machine' as Agent can serve to avoid regarding (17) (b) as a violation of Fillmore's principle according to which "only noun phrases representing the same case are conjoinable" (Fillmore 1968a 22) : "either he or a machine will do". Lastly, I suggest

that Langendoen's <u>Cause</u> "role" is not eligible to represent '<u>machine</u>' in the sentences of (17); this new role or case was set up by Langendoen to <u>contrast</u> with the Agent role, and what we need here, as I have argued, is just the Agent case, to represent the fact that a 'personification' of '<u>machine</u>' is compatible with the beliefs that some people hold.

I now wish to turn to a second, related, type of situation. So far, my analyses have proceeded under the assumption that the grammatical object of the verb '<u>use</u>'(or the corresponding <u>of</u>-phrase governed by the noun '<u>use</u>') are to be analyzed as Instruments; this is the generally accepted analysis in TG literature (Lakoff 1968b, Fillmore 1968a, 25). But surely, '<u>use</u>' can have the sense of '<u>employ</u>' as in "Our department uses <u>two typists</u>", and the sense of "behave towards" as in "Use <u>others</u> as you would like them to use <u>you</u>" (Hornby). In addition, consider the following example :

(18)

The <u>use</u> of <u>machines</u> instead of <u>human beings</u> to do the intellectual tasks of <u>abstracting</u> and <u>indexing</u> (J 471).

It seems to me that the reasoning adopted above for "thinking machines" is applicable here. Provision needs to be made for '<u>human beings</u>', '<u>men</u>', '<u>man</u>', etc. to be marked [- Animate] in order to be eligible for selection as Instrument with verbs like '<u>use</u>'and cognate nominals. The only difficulty which I can see about such an account is that, at least in my speech, it is impossible to say "x The application of machines instead of human beings to abstracting and indexing", although it seems that '<u>machines</u>' and '<u>human beings</u>' in this example should also be analyzed as Instrument. At present I can see no way of resolving this difficulty.

The next problematic example I wish to discuss is given in (19) (a). Note that '**select**' and '**selection**' do pass tests (1) and (11) : 'we use/apply machines to select information'. I regard (a), (b), and (c) as equivalent for the purposes of this discussion.

(19)

(a) Sentence selection <u>by men and machines</u> (J 816).
(b) Sentences selected by men and machines.
(c) Men and machines select sentences.

What I wish to argue about (19) is this. I believe that Fillmore's principle that only noun phrases representing the same case are conjoinable correctly predicts that (19) (a), (b) and (c) are unacceptable under <u>one</u> interpretation, namely that corresponding to just "men select sentence by (or with) machines". As the full abstract shows, however, what we have here is   a <u>comparison</u> between "sentence selection by men" and "sentence selection by machine (methods)" :

(20)

The formation of abstracts by the selection of sentences. Part I. <u>Sentence Selection by Men and Machines</u>. Auto-abstracting techniques based on high-frequency words show an extremely small variation among themselves in the selection of sentences to form abstracts. Human selection of sentences, although less variable than chance expectancy, is considerably more variable than the machine methods. There was very little agreement between the subjects and machine methods in their selection of representative sentences. Author.
Part II. The Reliability of People in Selecting Sentences. Although subjects exhibited some reliability in selecting representative sentences, the resultant reliability was low. The lack of inter- and intra-subject reliability seems to imply that a single set of representative sentences does not exist for an article. It may be that there are many equally representative sets of sentences which exist for any given article. (J 816).

We must therefore assume that "sentence selection by men and machines" in this abstract has an underlying structure consisting of two sentences or an equivalent construction with two nominals : something like "sentence selection by men (= A) and sentence selection by (means of) machines (or 'machine methods' (= I) COMPARED".

The last point I wish to examine in this section concerns the prepositional phrases underlined in the following quotations; the predicates which I assume to govern these prepositional phrases are capitalized :

<div align="center">(21)</div>

(a) machines are powerful tools which, <u>under the guidance of human judgment</u>, can DO routine operations that free the creative talents of man (J 436).

(b) the system would search a million documents at a cost of (...), which he compares with (...) as an estimate of the cost of the same WORK <u>with an entirely human agency</u> (J 689).

(c) These machines are USED one after the other, automatically <u>without human intervention</u>, each machine performing efficiently one of the stages in the operation (J 684).

To my knowledge, such 'human' prepositional phrases have never been discussed within case theory, at least in clearly active sentences like (a). Nor has the problem of prepositional phrases with '<u>without</u>', as in (21) (c). I shall not attempt to examine (21) (b); about this sentence, I would just say that for some reason that no doubt has to do with "<u>entirely</u>", and Instrument is not possible in it : "x the same work done by machine with an entirely human agency". Nor shall I attempt to analyze the other two prepositional phrases internally. All I am concerned with is their case status in the sentences just quoted. In (22) and (23) below I give paraphrases of (21) (a) and (c). These paraphrases suggest that the prepositional phrases under discussion are

to be analyzed as Agents. Note in particular that in (22) (a),
(b), and (c), "we" or "us" and "under the guidance of human
judgment" are redundant, and that (23) (a) and (b) are also
dubious for the same reason. Note, furthermore, that (22)
(d) is acceptable because, unlike (22) (a), (b), and (c), it
does not violate Fillmore's principle that "each case relation-
ship occurs only once in a simple sentence" (Fillmore 1968a, 21).

<div align="center">(22)</div>

(a) x machines are powerful tools which, under the
guidance of human judgement, we can use to do
routine operations...

(b) x machines are powerful tools with which we can do
routine operations under the guidance of human
judgement.

(c) x machines are powerful tools which, under the guidance
of human judgement can be used by us (documentalists)
to do routine operations ...

(d) machines are powerful tools which, under the guidance
of human judgements, can be used to do routine
operations...

(e) machines are powerful tools which we can use (can be
used) to do routine operations under the guidance of
human judgement.

<div align="center">(23)</div>

(a) ? They use these machines one after the other without
human intervention.

(b) ? The machines are used by them (these documentalists)
one after the other without human intervention.

(c) They use these machines one after the other to
retrieve information without human intervention.

If my analysis of these prepositional phrases as Agents is
correct, an important conclusion can be drawn from (22)
(e) and (23) (c): sentences of the type "x uses y to
do something" and "x does something with y" are not the same.
More precisely, sentences of the former type cannot be regarded
as "simple sentences" in Fillmore's sense, and those of the
latter type can.

I shall end this section by returning to another important conclusion drawn in it, and by suggesting that it could be extended to other word classes than those I have discussed. I have argued that for certain words the theory should allow for mutually incompatible lexical entries, corresponding to differing beliefs. The conclusion is not merely important as a general theoretical point. It is also important from the more technical point of view of case theory, because one of the claims originally made about this theory was precisely that it <u>simplified</u> lexical entries of the type proposed in Chomsky 1965 (see, in particular, Fillmore's paper on 'hitting' and 'breaking' : 1967).

As Fillmore realizes, another class of words which present problems for case theory is what he calls 'human institution' nouns (Fillmore 1968, 24, fn. 31). The suggestion on which I wish to end is that such nouns also require multiple lexical entries, perhaps via features such as [+ Animate] and [- Animate]. To take the words '<u>center</u>' and '<u>library</u>' as examples, I suggest that they may   be analyzed as Agent in (24) (25),and as Instrument in (26). Also compare the examples of (26) with a much more complex situation, in (27).

(24) At the time of this report, the <u>Center</u> was <u>planning</u> to change over to a mechanized <u>storage</u> and <u>retrieval</u> system ... (J 1).

(25) Cp. : <u>They</u> (the people) were planning/The Center <u>were</u> planning

(26)

(a) A system of information systems. Proposes aid to the
information crisis through the establishment of a
network of information centers... (J 146).
(b) ... in order to increase the holdings at the Center
and thereby make it a more effective tool in research
(J 127).
(c) Twelve functions of a library viewed as a system of
communication are listed (J 599).
(d) ... a pilot center for communicating information...
(J 203).

(27)

Directory of R and D information systems;
a listing of centers, services, sources and
systems engaged in collecting, storing and
disseminating scientific data and information
applicable to aerospace research and technology.

This volume presents the findings of a survey of in-
formation centers, services, sources and systems
collecting, storing and distributing information and
data useful in Air Force research and development.
It includes descriptions of content and services of
each system, and brief instructions for their use.
Various search methods are provided for locating
information sources pertaining to narrow or broad
areas of interest. (J 994).

3.5.3. 'Use' and other instrumental predicates.

For the purposes of this discussion, it will be sufficient
to define an instrumental predicate as one whose grammatical
object in active sentences with this predicate as main verb
can be analyzed as an Instrument, or as any predicate seman-
tically relatable to the former type. In the literature with
which I am acquainted, the only instrumental predicate whose
existence is recognized and to which some attention has been
devoted is the verb 'use'. 'Use', however, is by no means
the only instrumental predicate. Among other English
instrumental predicates I would mention the following, which
I have found in the Janaske corpus :

(28)

```
apply I
I be available
  be based (up)on I
  depend (up)on I
  derive from I
  draw (up)on I (from I)
  effect, effectiveness of I
  employ I
I go far towards
  importance of I
I have light to shed on
I make (possible, etc.)
  is a matter of I
  operate I
I operate
  refer to I
  resort to I
  start from I
  utilize I
I be a way to
```

Here, however, I shall not attempt to set up a complete inventory or a detailed analysis of such verbs because I wish to concentrate on what I believe to be a more important problem, namely special properties of USE (and of synonyms like 'employ') which, to my knowledge, have never been discussed.

First, USE accepts what may be called a compound, or double Instrument, which is partly similar to Fillmore's "car's fender" example. As far as I can see, all the sentences given in (29) are acceptable (except when otherwise specified) and they are semantically equivalent, except for the presence or absence of an Agent. Moreover, "the machine" can be regarded as the primary Instrument and "punched-cards" as the secondary Instrument. Corresponding, attested, examples are given in (30). A further possibility of paraphrase, namely 'IN the machine, they use punched-cards', will be discussed in 3.5.5.

(29)

(a) They USE the machine and punched-cards.
(b) They USE the machine with punched-cards.
(c) They USE punched-cards with the machine.
(d) With the machine they USE punched-cards.
(e) The machine USES punched-cards.
(f) x Punched-cards USE the machine.
(g) x With punched-cards they USE the machine.
(h) Punched-cards are USED with the machine.
(i) The machine is USED with punched-cards.

(30)

(a) A concept of documentation ... BASED upon the following procedures (J 851).
(b) The system described EMPLOYS three established techniques (J 35).
(c) The system USES three types of punched-cards (J 19).
(d) They [= punched-cards] can be USED with or without machines (J 328).

Second, there are good reasons to believe that certain verbs - for instance present - do not have Instruments in their case frame. Compare 'The key open e door', 'This example suggests that' and 'x Various ex les present the method', as against 'Various examplas serve to/are used to/ etc. present the method'. This throws light on an important property of instrumental predicates like USE. They provide a way of introducing simple Instruments into sentences whose main verbs do not accept such Instruments. This is particularly clear with some verbs of the type discussed in 3.1. Consider the following examples :

(31)

(a) Based on the idea that..., he discusses... (J 290).
(a') x The idea discusses...
(b) In addition to defining... by using a simple illustration of..., the author shows... (J 82).
(b') x An illustration defines...
(c) Proceeding from this, the author elaborates with... (J 199).
(c') x This elaborates with ...
(d) Based on the author's findings the report recommends... (J 980).
(d') x The author's findings recommend.

Incidentally, note that the above examples also suggest that certain prepositional phrases such as are exemplified below in (32) could have 'based on' or the like as their underlying structure. Note, in particular, the use of the preposition 'on'. Lastly, note that such expressions are an important mark of discourse structure, since they can refer to all that precedes in a given text.

(32)

(a) On these bases, the paper presents... (J 963).
(a') x These bases present ...
(b) Upon this foundation, he criticizes ... (J 54).
(b') x This foundation criticizes...
(c) On the premise that—, the writer discusses... (J 289).
(d) x The premise that— discusses...

Another remarkable property of verbs like USE is that
simple sentences containing just an instrumental predicate
are somehow incomplete : one always uses something TO DO
SOMETHING. Now, as already noted, purposive infinitival
clauses also occur with non-instrumental predicates. Compare
(33) and (34) :

<div align="center">(33)</div>

(a) The author presents the method USING various
    examples (J 330).
(b) The author USES various examples to present
    the method.

<div align="center">(34)</div>

(a) The author GIVES various examples to present
    the method.
(b) The author presents the method by GIVING various
    examples.

I suggest that a semantic representation of the sentences in
(33) and (34) must capture the similarity between them and
that this can be done along the lines proposed above in 3.4.,
as diagrammed below in (35) and (36).

<div align="center">(35)</div>

(a)

(b)

(36)

(a)

```
                              SO -
            Cause                        Goal
            + Means                      + Purpose
              |                            |
             S1+                          S2-
        V     A      Topic ?
        |     |        |
    present author examples       to present the method


                        SO-
            Cause                    Goal
            + Means                  - Purpose
              |                        |
             S1-                      S2+

        by giving examples       the author presents the method
```
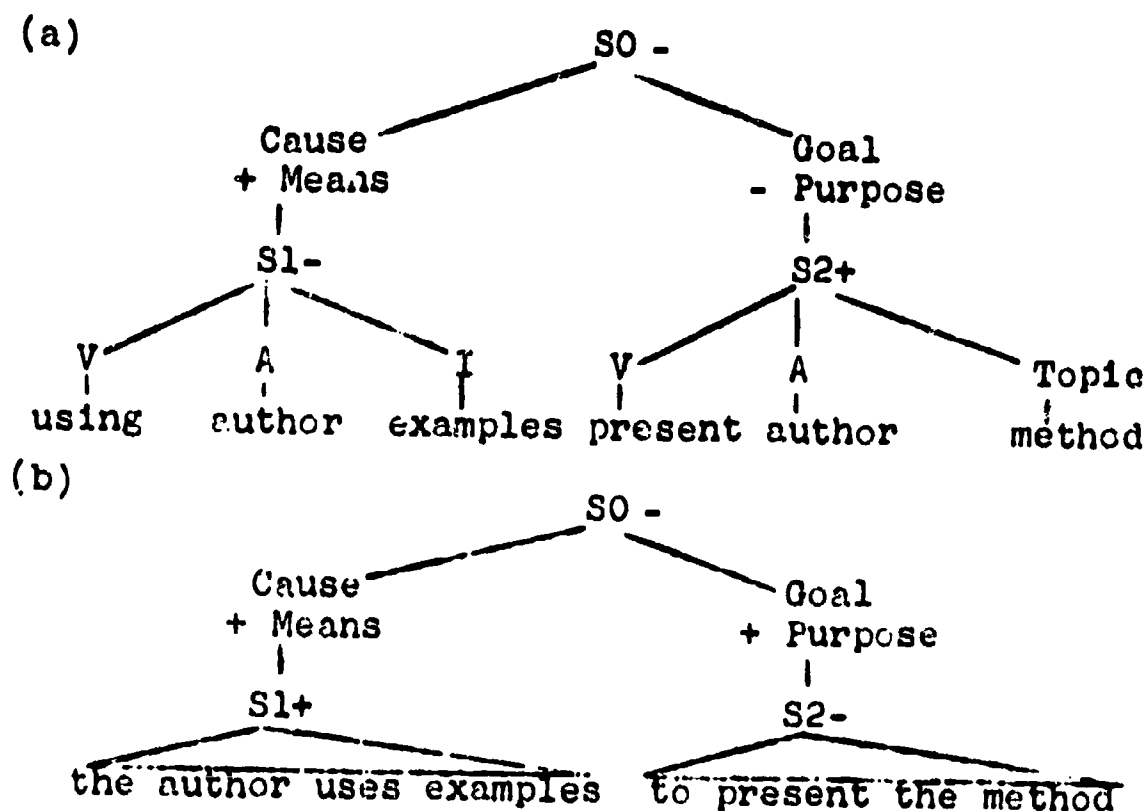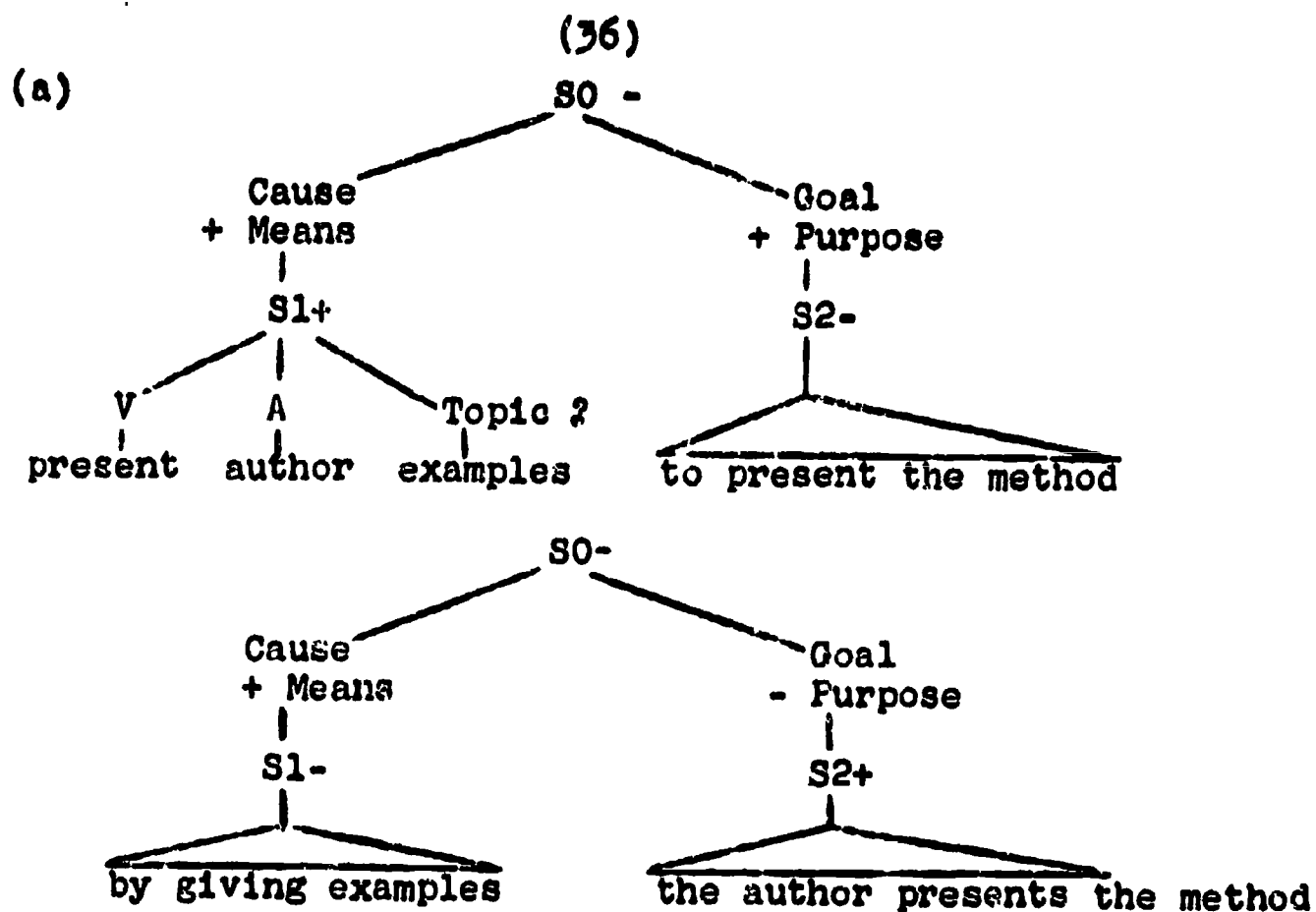
The discussion of the similarities and differences between sentences of the types illustrated in (33) and (34) can be carried a step further by considering the following situations :

	Cause-clause	Goal-clause	Examples
(a)	DESIGN etc. + Goal	I=Goal of Cause-clause	They have designed a machine to retrieve information (with).
(b)	USE + I	I=I of Cause-clause	They use a machine to retrieve information (with).
(c)	(a) + (b)	(a) + (b)	a machine designed and used (or: to be used) to retrieve information (with).
(d)	USE + I	I different from I in Cause-clause but relatable to it	they use a machine to retrieve information with punched-cards.
(e)	USE + I	I different from, and unrelated to I in Cause-clause	the number of matching terms between the list of request terms and the terms used in indexing a document is used to select the documents by computer.
(f)	V + Object	I=Object of Cause-clause	By throwing a stone (at it) he broke the window.
(g)	V + Object which cannot serve as I of Goal-clause together with the I of this clause		He climbed the tree to watch birds with his field-glasses.
(h)	negative		select documents (by sight) without the use of needles

(37) (b) - and, in particular, the possibility of an anaphoric 'with' in the to-clause - suggests that a simple Instrumental or a Cause can be posited in the Goal-clause whenever this is allowed by its main verb. Compare (38) (a) and (b) :

(38)

(a) They use a machine to retrieve information (with).
(b) x They use an example to present the method with.

If this is correct it follows that sentences of the type 'x uses y to V something' and those of the type 'x V's something with y' are equivalent except if the V does not accept a simple I. If it does , the use-clauses can be regarded as a device for 'factoring out' the Instrument and for avoiding its repetition. Compare the attested example in (39) (a) with its paraphrase in (b) :

(39)

(a) the libraries will use machines for cataloguing, indexing, abstracting, cross-referencing, and retrieving of information (J 363).
(b) the libraries will catalogue, index, abstract, cross-reference and retrieve information with machines.

I suggest that (37) (a), (b) and (c) can be analyzed along the lines of my proposals about asymmetric conjunction. Indeed, we have at least the possibilities given below in (40) (a) and (b), which may be compared with the attested examples in (40) (c) and (d). The analyses proposed for (37) (a), (b) and (c) are diagrammed below in (41) :

(40)

(a) the machine was designed AND is used to search information.

(b) the machine was designed FOR use in searching information.

(c) an automatic microfilm searching machine designed to increase searching speed, to edit out extraneous matter, and to provide copies of the desired information (J 1109 : the to-clauses seem to be conjoined symmetrically).

(d) Systems and procedures developed for the search, coding and mechanized processing of bibliographic information on thermophysical properties.(title) Methods used by ... for handling collected bibliographical information on thermophysical properties are described (sentence 1, J 985).

(41)

(a)


```
                        SO -
           Cause                    ·Goal
                                    + Purpose
             S1+                      S2-
         V        Goal          V       I      O
                                                NP
    was designed  the machine  to retrieve the
                                        machine information
```

(b)


```
                        SO -
         Cause                      Goal
                                    + Purpose
           S1+                        S2-
        V        I                 V     I      O
    is used  the machine      to retrieve the  (for)
                                       machine  information
```

(c)


```
                              SO -
                  Cause                    Goal
                    |                      + Purpose
                   Sl+                        |
            Cause         Goal               S2-
              |           +Time sequence
             Sl'+         S2'-
        The machine was  to be used      to retrieve information
        designed
```
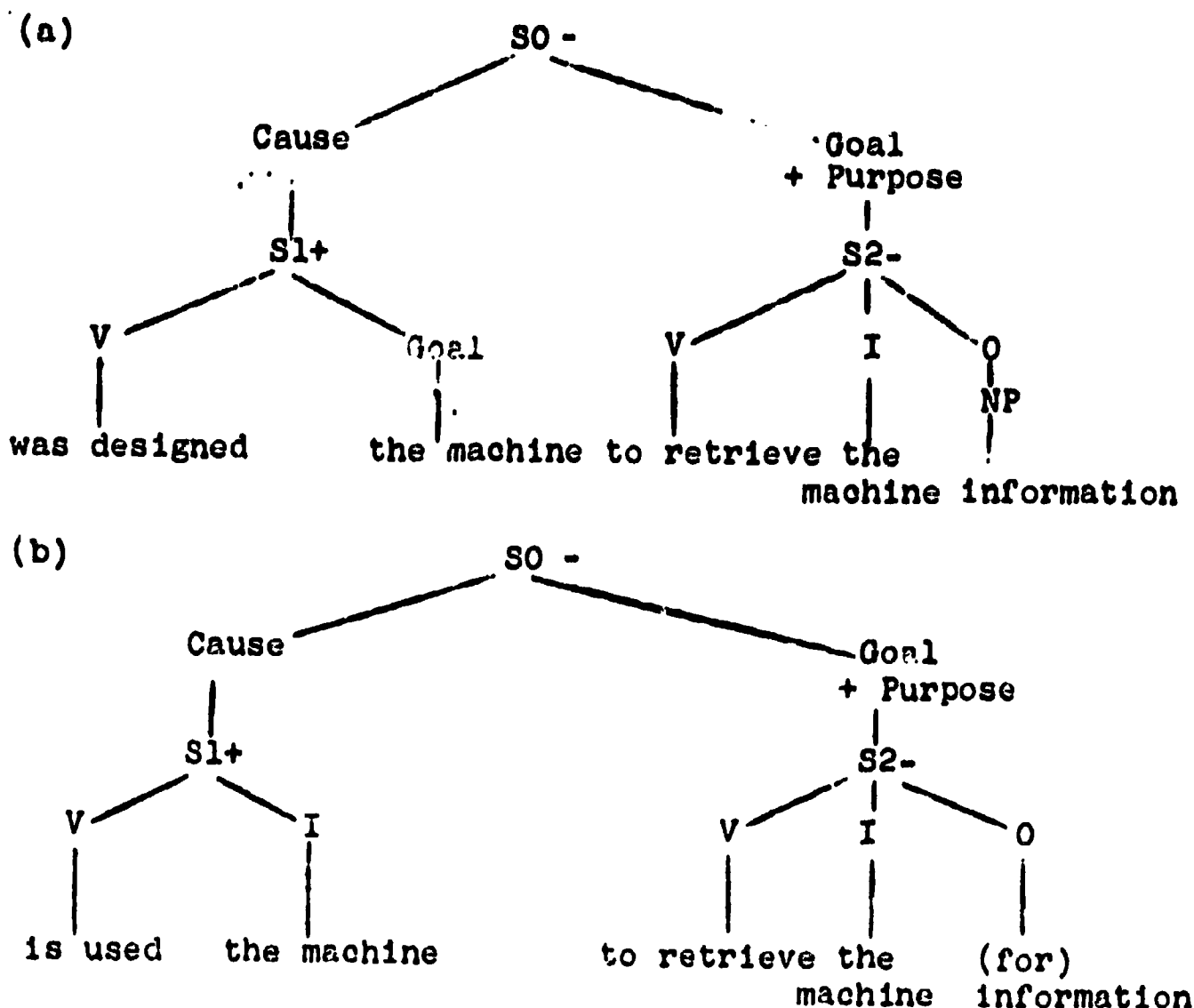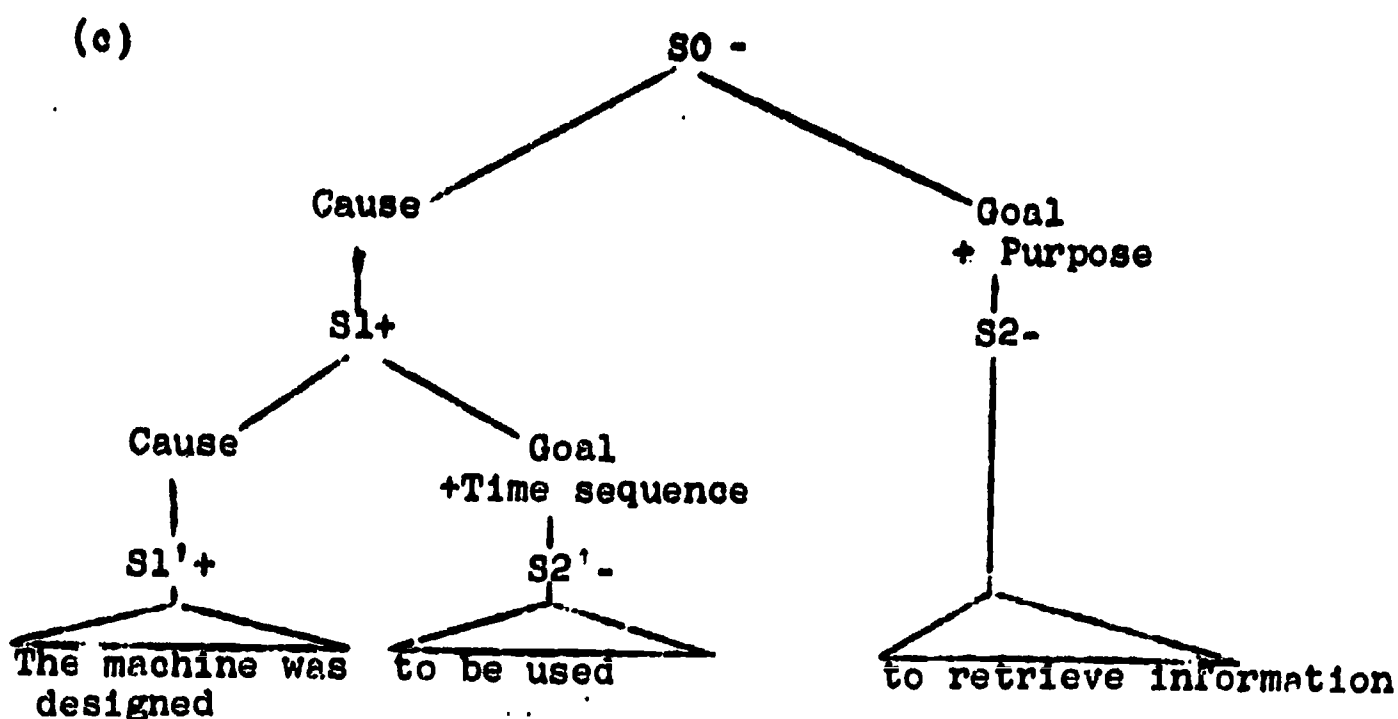
A comparison between (37) (d) and (e) suggests that one of the functions of Cause-clauses with '<u>use</u>' and equivalent predicates is to permit the introduction of mor<u>e</u> than one Instrument in the same sentence. When these Instruments are related to each other and can be conjoined - as in (37) (d) - the sentences with a <u>use</u>-clause are equivalent to the corresponding sentences with instrumental prepositional phrases, as is illustrated in (42). The other type of situation is illustrated in (43) :

(42)

(a) They use a machine to retrieve information with punched-cards.
(b) They retrieve information with a machine and with punched-cards.

(43)

(a) The number of matching terms between the list of request terms and the terms used in indexing a document serves to (is used to) select the documents by computer.
(b) x The number of matching terms and the computer select the documents.
(c) x The number of matching terms and the (a) computer serve (are used) to select the documents.
(d) x They select documents by computer and by the number of matching terms...

A comparison between (37) (f) and (g) suggests that various sentences are convertible into 'instrumental sentences'. When the Object of the Cause-clause is eligible as Instrument of the Goal clause, the result of the conversion can be a 'simple' instrumental sentence, as is shown in (44). Otherwise, but provided the Instrument of the Goal clause is relatable to the Object of the Cause-clause, the sentence can be converted into a sentence with 'use'; this is shown in (45) and in the attested example of (46):

(44)

(a) By throwing at stone (at it), he broke the window.
(b) He broke the window with a stone.
(c) He used a stone to break the window.

(45)

(a) He climbed the tree to watch birds with his field-glasses.
(b) By climbing the tree, he watched birds with his field-glasses.
(c) x He watched birds with the tree and his field-glasses.
(d) He USED the tree to watch birds with his field-glasses.

(46)

(a) By noting the number of matching terms between this extended list of request terms and the terms used in indexing a document (= 0 of 'noting'), and with due regard for their degree of association (=0 of 'regard"; compare "considering", etc.), documents are selected by the computer (=I) (J 942).
(b) They USE the number of matching terms and their degree of association to select documents by computer.

Attested examples of negative Cause-clauses are given below in (47). Note that in (47) (b), the by- and the without phrase are conjoinable as Instruments. The advantage, however, of analyzing the without-phrase as a Cause-clause is that such an analysis captures the similarity between this without-phrase and the "even though" clause in (a) :

(47)

(a) <u>(a system) which</u> (=I) can find documents related to a request <u>even though they may not be indexed by the exact terms of the request</u> (=C=.without their being...
: J 942).

(b) the card selection can be done <u>by sight</u> (=I) <u>without the use of needles or other mechanical selection apparatus</u> (= C = without using...).

3.5.4. The instrumental role of 'system' and related words.

A striking feature of the corpus under investigation is
the part played by words like 'system' in allowing for the
expression of conceptually complex instrumental relationships.
In order to throw some light on the question, I shall posit
that the sentence types given below in (48), exemplified in
(49), and symbolized more abstractly by tree diagrams in
(50) represent the main types of Instrumental relationships
which 'system' and related words can contract. In the formulas
of (48) and in the trees of (50), the function of the NP
with 'system' is represented in main clauses even though the
word 'system' or related words occur with relative clauses
or relatable constructions in some corresponding examples of
(49). Note that in (48) and (49) 'system' is or can be the
grammatical subject in all the examples, except in (f) and (g)
where such a grammatical function is ruled out for 'system' as
Instrument. (48) and (49) (a) represent a sentence type in
which it seems that 'system' enters into an Instrumental rela-
tionship with a (let us say) 'broader' Instrument than itself,
for instance 'theory'. In (48) and (49) (c), on the contrary,
'system' seems to contract a relationship with a 'lesser'
Instrument.

(48)

(a) system (= I) + is BASED on + theory (= I).
(b) system (= I) + is USED + TO + V (or : FOR Ving).
(c) system (= I) + USES + technique (or machine, etc.= I).
(d) system (= I) + is BASED on (DEPENDS on) + Ving.
(e) system (= I) + non-instrumental Verb + I.
(f) Verb (not accepting 'system' as subject) + system (= I?).
(g) Verb (not accepting 'system' as subject) + system (= I?).
+ TO + V (or : FOR Ving).
(h) =(g) but with Verb accepting 'system' as subject.

[though perhaps ... relatable words like 'method'
and 'technique' .]

(49)

(a) he criticizes six <u>theories upon which</u> experimental retrieval <u>systems have been based</u> (J 54).
(b) <u>Methods used</u> (by ...) <u>for handling</u> collected bibliographical information ... (J 985).
(c) <u>The system</u> described <u>employs</u> three established <u>techniques</u> (J 35).
(d) Most mechanized information retrieval <u>systems depend upon extracting</u> informatior from documents...(J 321).
(e) <u>This system</u> handles large numbers of chemical compounds and polymers <u>by their structural and compositional features</u> (J 275).
(f) It should be possible to set up, in the near future, a <u>procedure</u> (= system) <u>whereby</u> one can either subscribe to scientific journals on a yearly basis or purchase single articles at a unit price (J 629).
(f') x The procedure can either subscribe... or purchase...
(g) a <u>technique</u> (= system) <u>for</u> literature indexing and searching <u>whereby</u> the notion of relevance is developed qu·ntitatively... (J 640).
(g') x the <u>technique</u> develops the notion of relevance quantitatively...
(g'')?<u>By (means of) this technique</u> the notion of relevance is developed quantitatively <u>for</u> literature indexing and searching .
(h) a <u>method for</u> arranging and finding documents <u>by assigning</u> to each what is basically a two-faceted, decimal/alphabetic classification designation (J 60).

(50)

(a)

+ instrumental   system   theory

501.

(b)

```
                              SO -
                    Cause              Goal
                      |               +Purpose
                      |                  |
                   S1+ (or:c)           S2-
            system ... used        to(for) Ving(N)
```

(c)

```
                           S
              V                          I
              |                    I            I
              |                    ¦            |
              |                    NP           NP
              |                    |            |
        +instrumental            system    system(or:machine,etc.)
```

(d)

```
                           S
              V                          I
              |                 I              I
              |                 ¦              |
              |                 NP           S(or:b)
              |                 |        ⎧ Ving  Case   Case...
        +instrumental         system    ⎩ N
```

(e)

```
                       S
        V      Case      Case              I
        |       /          \         I          +I
        |      /            \        ¦           _
        |    /               \       NP          NP
        |                             |           ¦
   -instrumental                    system
```

(f)

(g) (cp. b)

(h)
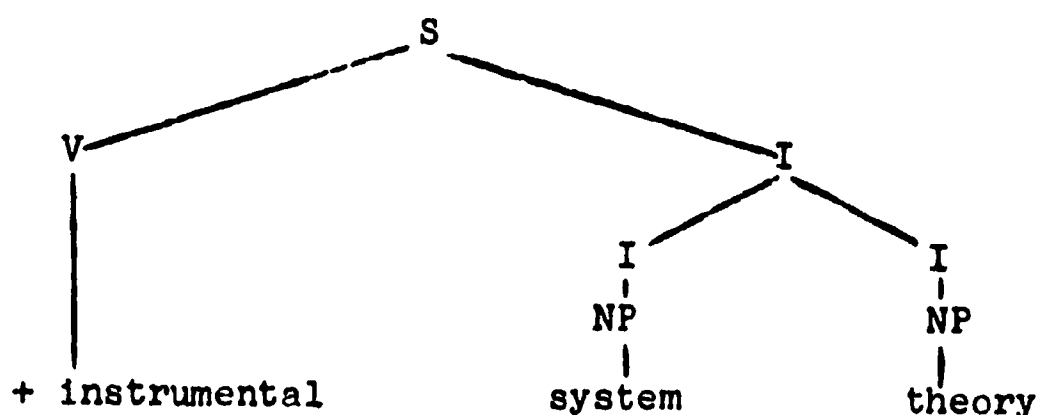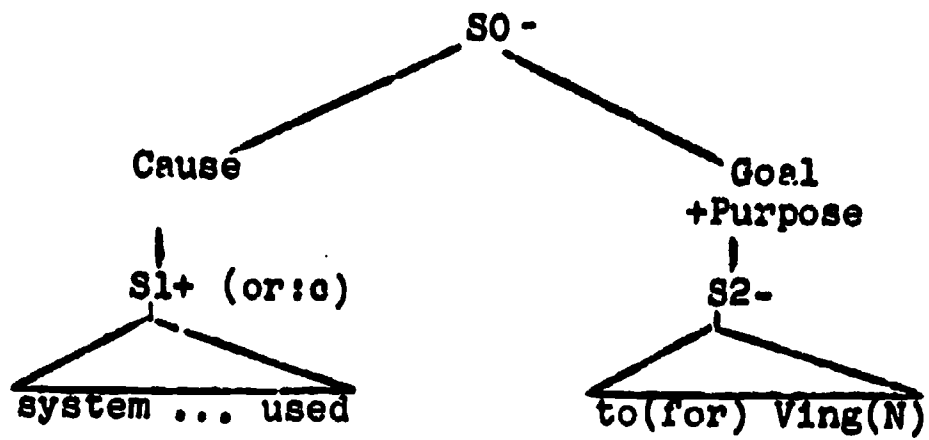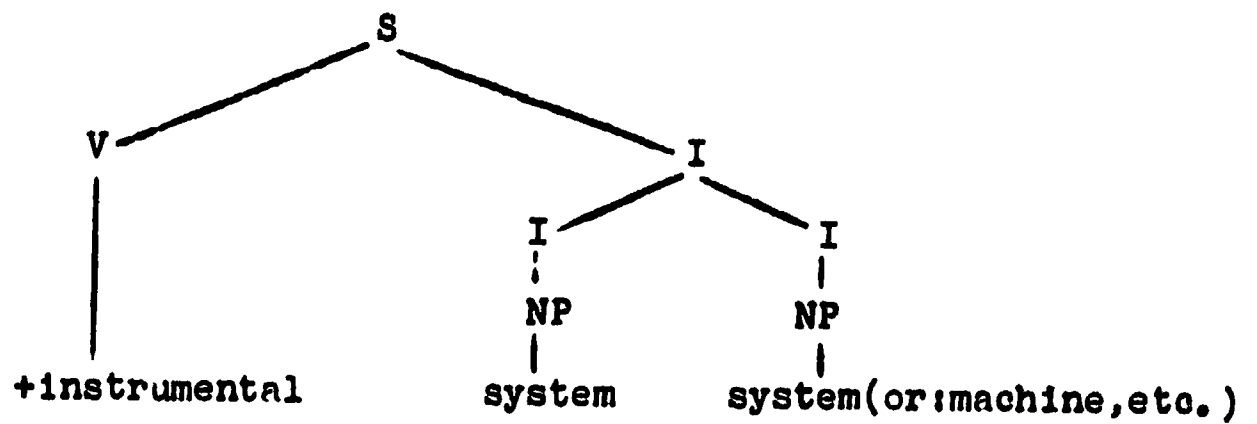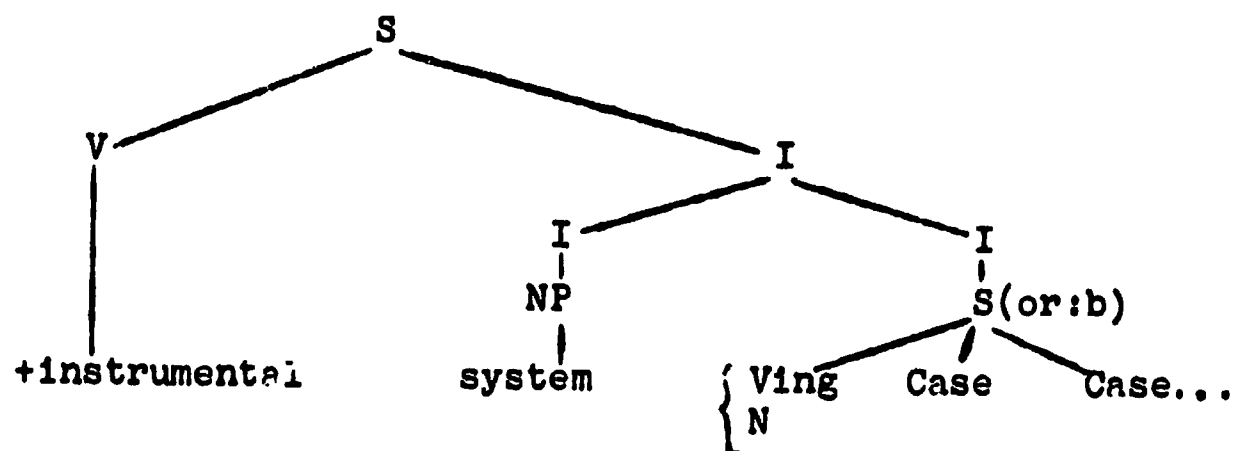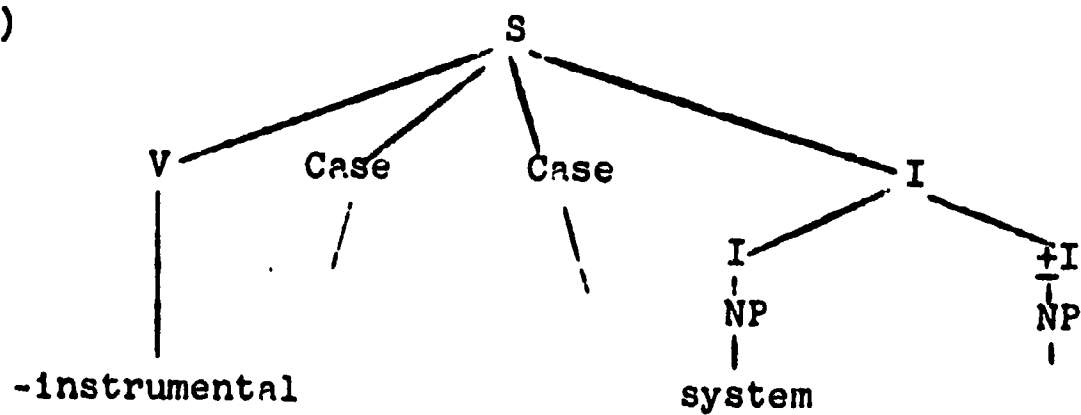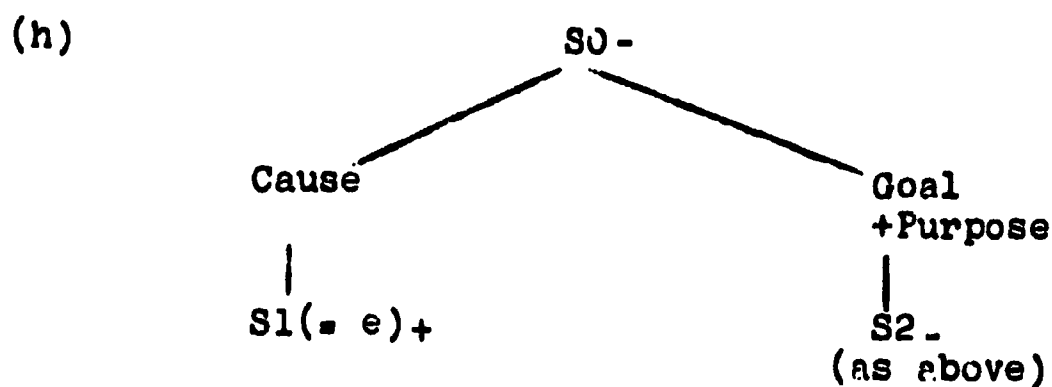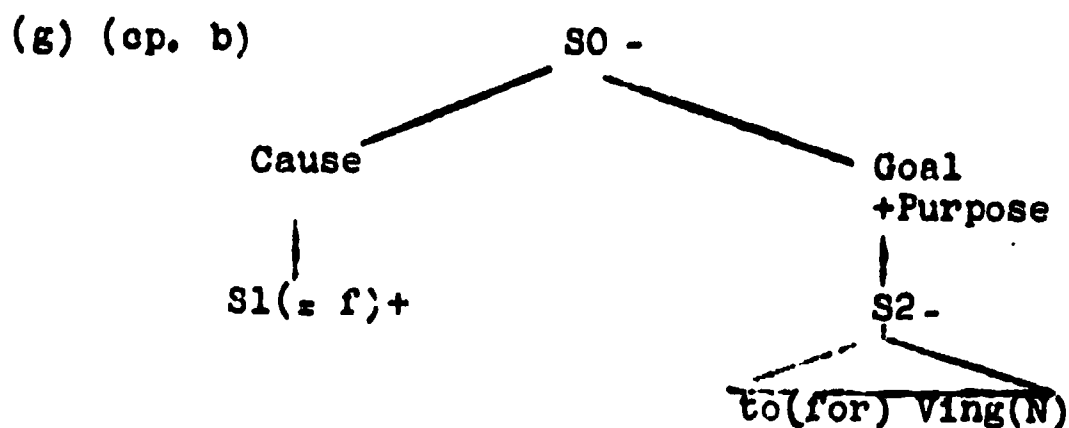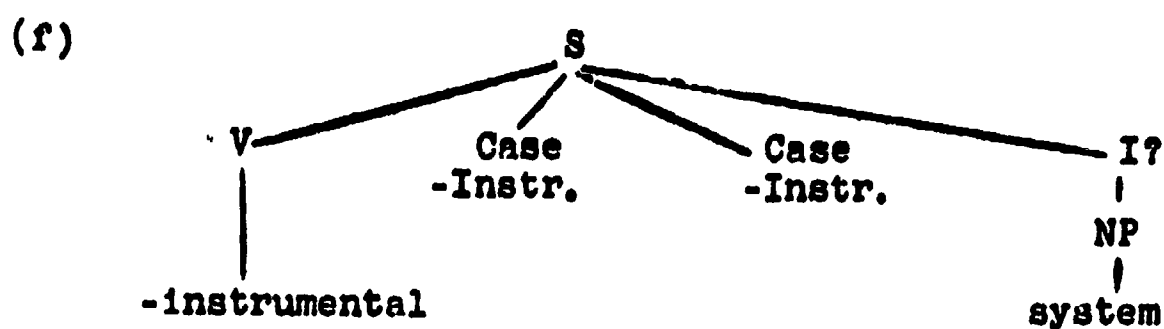
In the rest of this section, I shall present more detailed analyses which I propose for the sentences of (49). In addition, I shall give some nominal constructions which can be paraphrased after the model of these sentences and which, on this basis, can be assumed to have a deep structure similar or relatable to the sentences in question. In the examples, the irrelevant material is given between brackets, and the relative clauses are treated as main clauses, as I shall not discuss the problem of how to represent relative clauses. For some of the illustrative sentences, I have found no corresponding nominal construction.

(51)

(a) <u>Illustrative sentence</u> : (he criticizes six) theories upon which (experimental retrieval) systems have been based (J 54).

(b) <u>Proposed analysis</u> :

```
                        S
             V                      I
             |               I            I
             |               |            |
             |              NP           NP
             |               |            |
        be based on       systems      theory
```
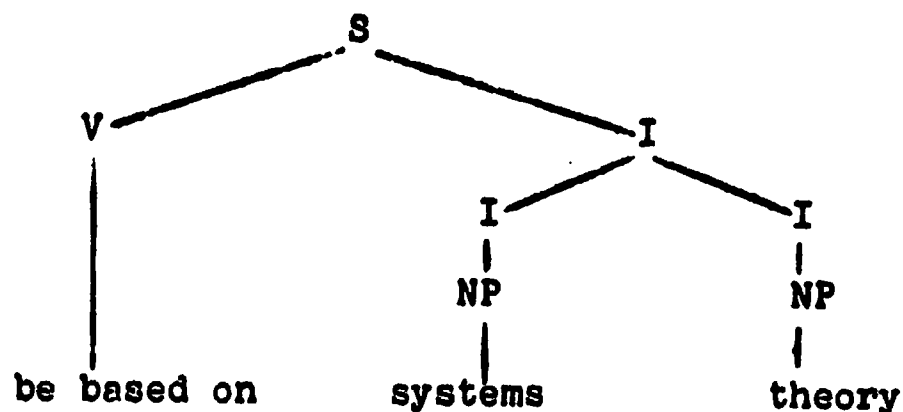
(c) <u>Corresponding nominal construction</u> : (retrieval) system theory (J 1044).

(52)

(a) <u>Illustrative sentence</u> : Methods used (by...) for handling collected bibliographical information ... (J 985).

(b) <u>Proposed analysis</u> :

```
                      SO -
         Cause                      Goal
                                  +Purpose
           |                        |
          S1+                      S2-
      V          I                  |
      |          |                  |
      |         NP              /        \
   is used    method        for handling...
```
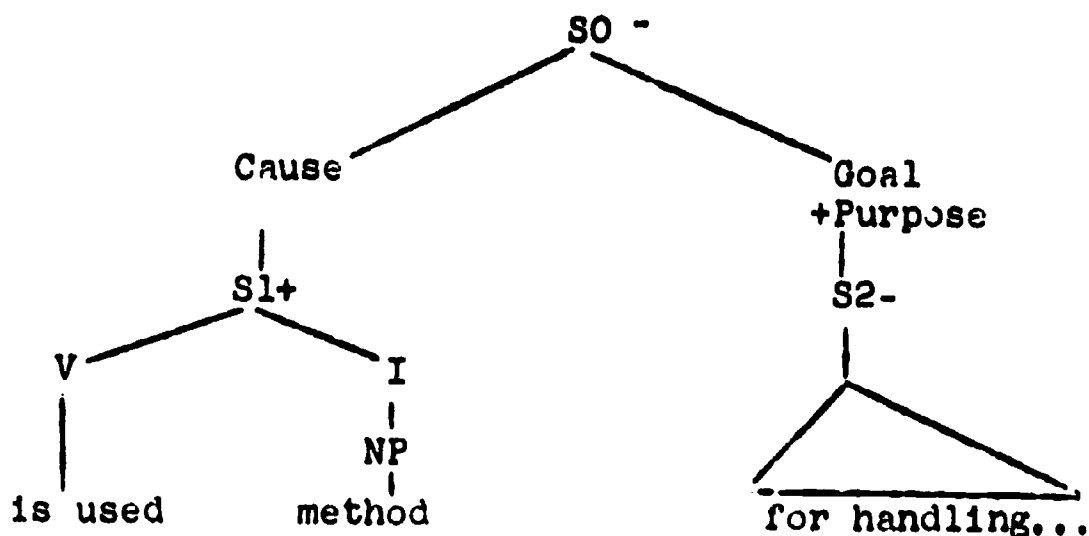
(c) <u>Corresponding nominal constructions</u>:
-information retrieval systems (J 321, etc.).
-a scheme for classification and coding of information (for machine retrieval) (J 580).

(53)

(a) <u>Illustrative sentence</u> : the system (described)
employs (three established) techniques (J 35).
(b) <u>Compare</u> : machine based systems (J 471)
(c) <u>Proposed analysis</u> :

```
                    S
            _____/_____
           V                   I
           |              ____/\____
           |             I           I
           |             |           |
           |             NP          NP
           |             |           |
        employs       techniques   system
```
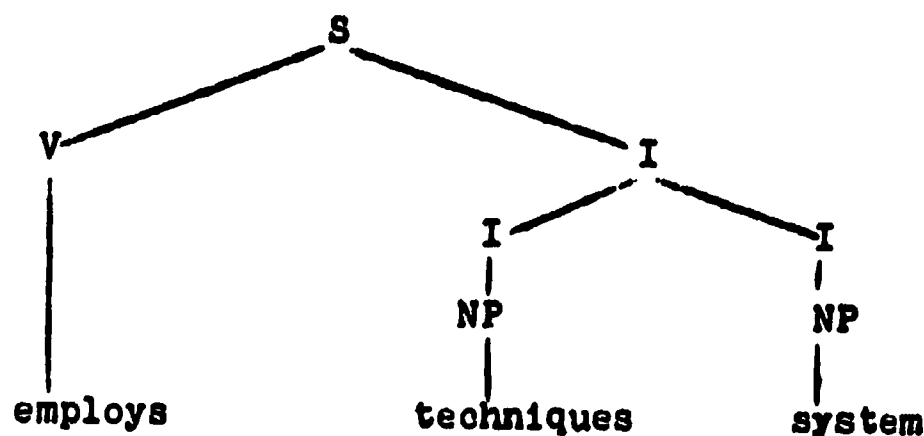
(d) <u>Corresponding nominal constructions</u> :
    - the punched card system (for indexing...) (J 830).
    - A system of information systems (J 146).

(54)

(a) <u>Illustrative sentence</u> : (Most mechanized information
retrieval) systems depend upon extracting information
from documents, (processing such information in some
manner, and finally placing this information in
a file so that the information can be retrieved
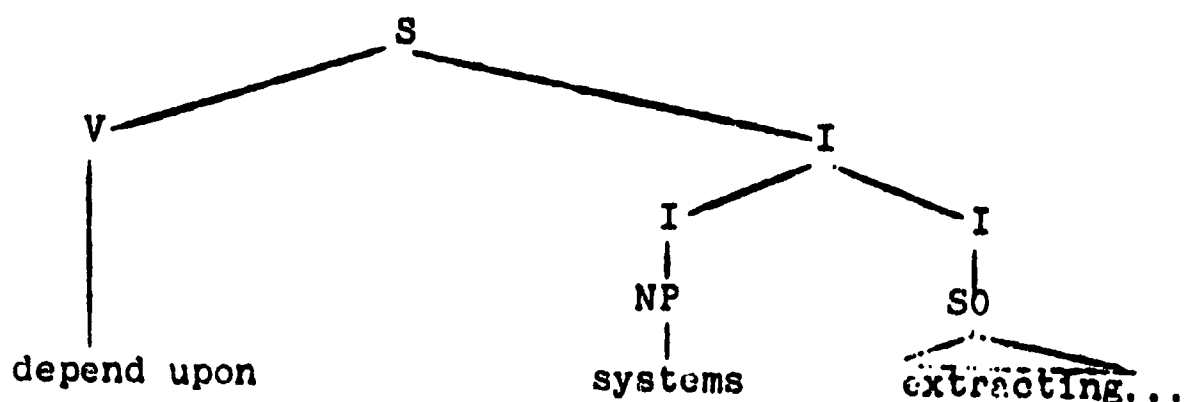by machine) (J 321).
(b) <u>Compare</u> :
    - (as long as searching) devices are based on
      the use of words (to describe documents) (J 804).
    - the system operates by feeding selected clue
      words into a computer which... (J 639).
(c) <u>Proposed analysis</u> :

```
                    S
            _____/_____
           V                   I
           |              ____/\____
           |             I           I
           |             |           |
           |             NP          S0
           |             |            /\
       depend upon     systems    extracting...
```

(d) <u>Corresponding nominal construction</u> : NONE.

(55)

(a) <u>Illustrative sentence</u> : This system handles large numbers of chemical compounds and polymers by their structural and compositional features rather than by their names (J 275).

(b) <u>Compare</u> : the (Peek-a-boo) system which achieves random access by means of superimposable cards (J 1104).

(c) <u>Proposed analysis</u> :
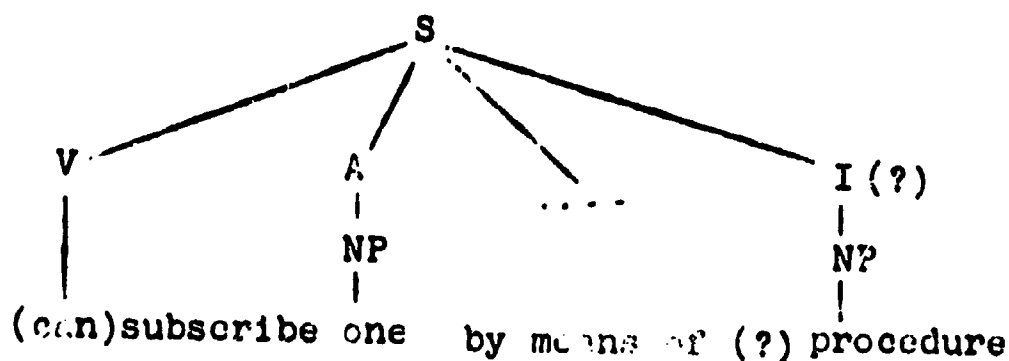
```
                    S
          _____/   |   _____
         V          O           I
         |          |        __/ \__
         |          NP      I        I
         |          |       |        |
         |          |       NP       NP
         |          |       |        |
      handles   compounds  system   by....
```
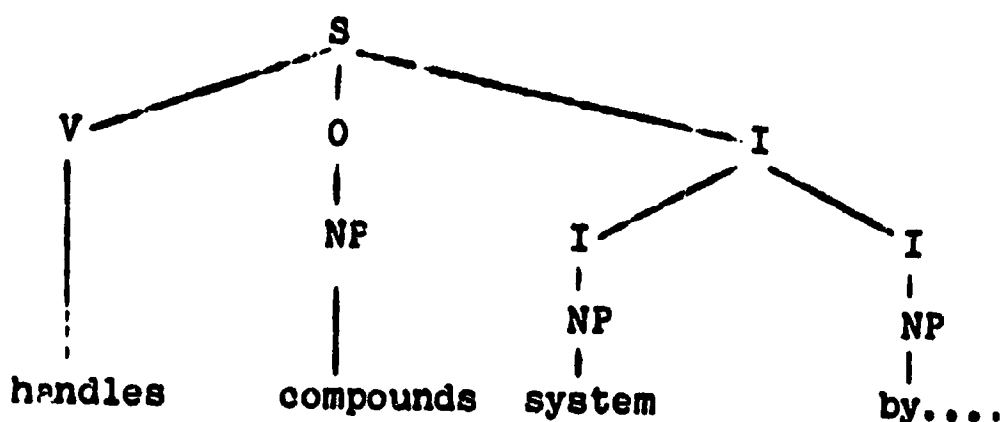
(d) <u>Corresponding nominal construction</u> : NONE.

(56)

(a) <u>Illustrative sentence</u> : (It should be possible to set up, in the near future) a procedure whereby one can either subscribe to scientific journals on a yearly basis ... (J 629).

(b) <u>Compare</u> : (Describes) the methods (developed at the Chemical Abstracts service) by which organic chemical data put on magnetic tape in notation form can be searched mechanically for specific, predetermined structural features (J 274).

(c) <u>Proposed analysis</u> :

```
                  S
        _____/  /  \   \_____
       V       A    ....      I (?)
       |       |               |
       |       NP              NP
       |       |               |
  (can)subscribe one   by means of (?) procedure
```
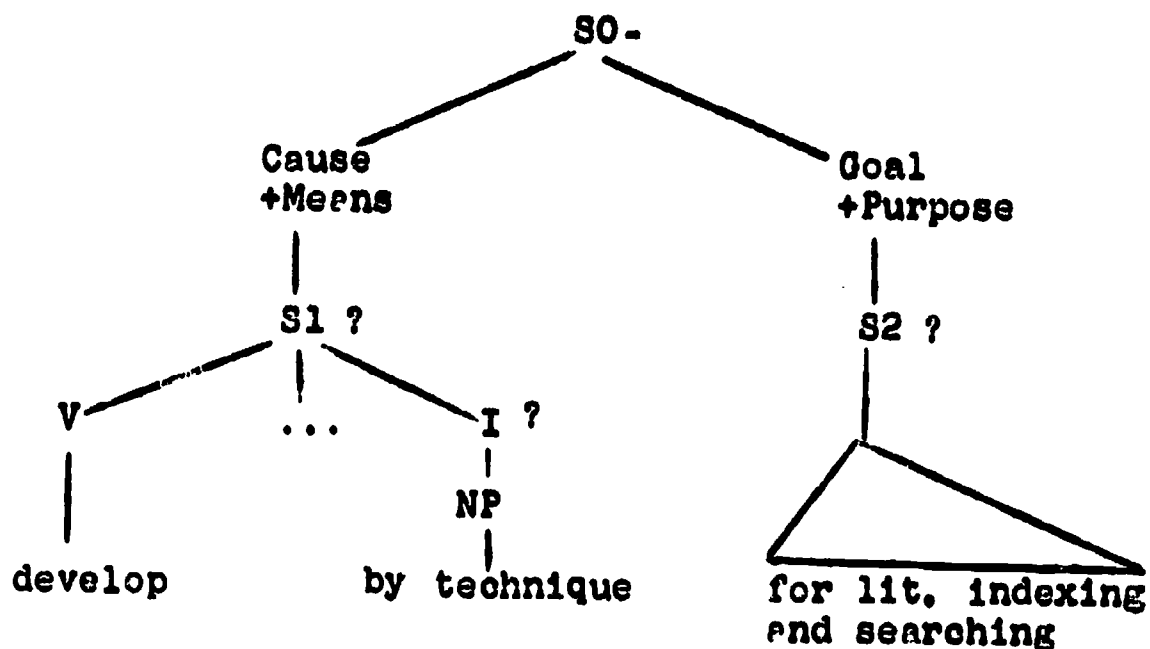
(d) <u>Corresponding nominal construction</u> : NONE.

(57)

(a) <u>Illustrative sentence</u> : a technique for literature indexing and searching whereby the notion of relevance is developed quantitatively... (J 640).
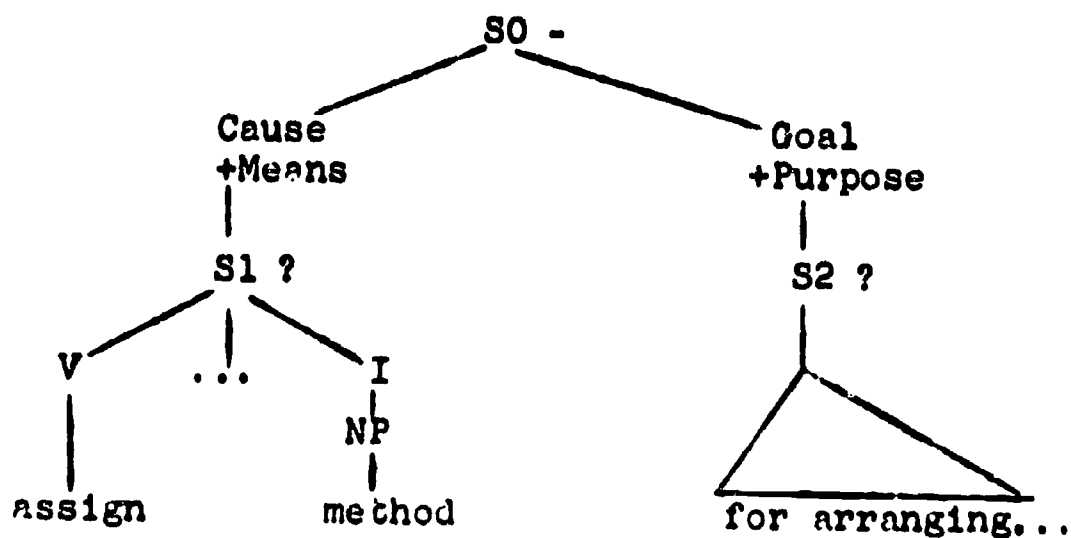
(b) <u>Proposed analysis</u> :

```
                        SO-
                   /          \
              Cause            Goal
             +Means           +Purpose
               |                |
              S1 ?             S2 ?
            /  |  \             /\
           V  ...  I ?         /  \
           |       |          /    \
           |      NP         /      \
        develop    |        / for lit. indexing
              by technique   and searching
```

(d) <u>Corresponding nominal construction</u> : NONE.

(58)

(a) <u>Illustrative sentence</u> : a method <u>for</u> arranging and finding documents <u>by</u> assigning to each what is basically a two-faceted, decimal alphabetic classification designation (J 60).

(b) <u>Proposed analysis</u> :

```
                     SO -
                 /          \
            Cause           Goal
           +Means          +Purpose
             |               |
            S1 ?            S2 ?
           / | \            /\
          V ... I          /  \
          |     |         /    \
          |    NP        / for arranging...
       assign  method
```

(d) <u>Corresponding nominal construction</u> : NONE.

Where a sentential paraphrase exists,   the nominal constructions could be analyzed in terms of the corresponding sentences. More precisely, they could be analyzed in terms of sentences with an abstract Instrumental verb which has no surface realization. For instance, "information retrieval system" would be analyzed as "system which Vinstrumental TO retrieve". In this fashion, the nominal constructions in question would be treated as instances of recoverable deletion.

Lastly, in order to out down the number of sentence types just presented, one may be tempted to do the following :
(1) type (f), namely "a procedure whereby..." could be reduced to type (d); indeed, "a procedure whereby..." and the like can be paraphrased as "a procedure (which) is BASED on ..." (op. "systems DEPEND upon Ving"). This analysis could also be adopted for the by-clauses of types (g) and (h).
(ii) The purpose-clauses in these sentence types could similarly be reduced to type (b); thus "a method for..." can be paraphrased as "a method (which) is USED for...", as has just been suggested for "retrieval system".
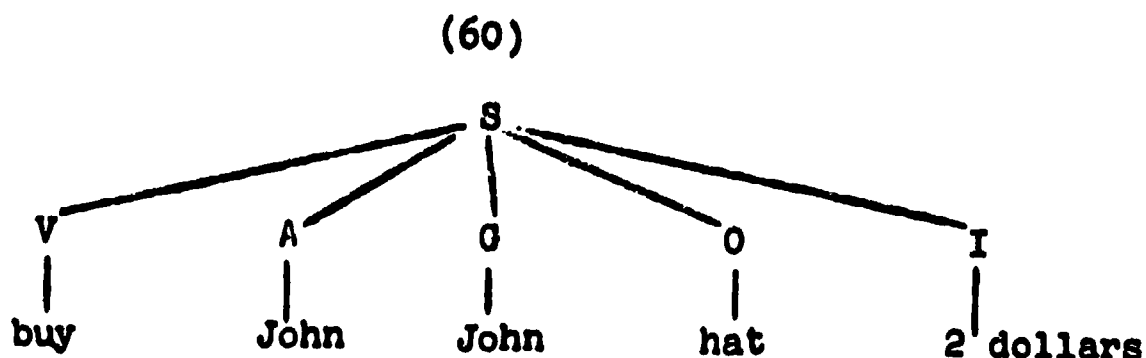
The last illustrative example, for instance, which is repeated as (59) (a), could ther be analyzed as in (59) (b), or as a corresponding relative-clause structure (59) (c) :

<div align="center">(59)</div>

   (a) A method for arranging and finding documents by assigning to each... a ... classification designation.

   (b) A method is used for arranging and finding documents. The method (it) operates (works) by assigning to each... a... classification designation. (is based on).

   (c) A method which is used for ... and operates by ... (is based on).

3.5.3. The locative role of 'system' and related words.

Fillmore (1969 a) allows for the possibility of represent-
ing a single surface constituent by two or more case nodes.
Thus, in "**John** bought a hat for two dollars" "**John**" would be
analyzed as both Goal and Agent :

(60)

```
                        S
        _____/|_____
       /        /        |        \         \
      V        A         G         O          I
      |        |         |         |          |
     buy      John      John      hat      2 dollars
```

The Goal-node in such a sentence usually has no surface
realization, and Fillmore (1970 b) proposes to delete it
transformationally. Note, however, that this is only usually
so : a salesman, for instance, can say "I shall buy it for
myself" (rather than for this firm : B. Mohan, personal
communication)

As far as I can see, the most important criterion for
setting up multiple case representations is the occurrence
of certain prepositional phrases and the use of certain
prepositions associated with a given item (e.g. "FOR myself"
in the above example). It is clear that the use of prepositions
is occasionally non-critical. For instance, by-phrases in
passive sentences neutralize Agent and Instrument : "done by
a machine/by people". Also consider the following attested
examples : "the Unit Card System for indexing" (J 757),
"the unit card system in the indexing of ..." (J 758), "the
revised unit card system of indexing" (J 758). Here, however,
I wish to present what I believe to be clear evidence for
a Locative analysis of words like 'system' and 'machine' in
sentences or cognate constructions in which they can also be

analyzed as Instruments. The evidence I wish to present is based on the use of certain prepositions. Consider the following examples, in which clearly Locative prepositions could be replaced by Instrumental prepositions such as 'by', 'with', or 'by means of' : (except perhaps with 'use', etc.) :

(61)

(a) cards are read on a microfilm viewer (J 779).

(b) Searches are now made on an ... computer (J 356).

(c) Of the two 0µtput systems designed, one stores the content of a selected microphotographs temporarily in an electrostatic tube and displays it on a cathode ray tube (J 37).

(d) a system by which most of the index of Chemical Abstracts could be coded mechanically on a computer for later searching on a computer (J 239, whose title reads : ... searching by the Western Reserve University searching selector).

(e) store and search information in a memory system (J 1113).

(f) system of dissemination of new scientific information in which machines are used (J 613).

(g) This system has been in use for about 10 years and has stored in it over 20,000 references (J 404).

(h) With proper manipulation of these associations (entirely within the machine) ... (J 942).

Also note that "The system uses three types of punched-cards" (J 19) has, next to various Instrumental paraphrases, already given, the following Locative paraphrases :

(62)

(a) In the system, they use three types of punched-cards.

(b) Three types of punched-cards are used in the system.

Lastly, let us consider the following examples in (63).
Two points can be made about the relationship between 'retrieve'
and 'machines' in these sentences. First, a case analysis is
necessary to throw light on what I have called asymmetric
conjunction : (63) (a), (b), and (c) make up a chain in which
(b) is a necessary condition for (c) and (a) a necessary
condition for (b). Second, (63) (c) and (d) suggest that
'machines' can contract both a Source and an Instrumental rela-
tion with 'retrieve' (on 'load', see Fillmore 1969a, 128).

(63)

(a) load information into the machines (= Goal)
    RESULTS IN load the machine (= Object) with information.
(b) information stored in the machines (= Locative).
    Cp. : the system has x stored in it (J 404).
(c) retrieve from the machines (= Source). Cp. :
    "librarians must not be unprepared to load the
    machines or to retrieve from them" (J 363).
(d) the machines (= Instrument) ... are only able to
    retrieve what has been put into them.

To conclude, I think that two consequences can be drawn
from the analyses outlined in this section. First, they suggest
that Fillmore's proposals for multiple case analyses are
essentially correct : I am not acquainted with any other theory
than his that would capture the facts about preposition
selection presented in this section.

Second, let us remember Fillmore's principle that "only
noun phrases representing the same case may be conjoined"
(Fillmore 1968, 22). While I have no attested example, it
seems to me that the following example is conceivable :

(64)

They retrieve information IN, WITH, and
FROM the computer.

If such a sentence is acceptable and if Fillmore's principle about conjoined NP's is to be preserved, it is necessary to reject Fillmore's proposal to have separate case nodes for co-referential entities (such as 'John' in 60). What seems to be needed instead is a feature system of the kind proposed in my discussion of asymmetric conjunction : a system in which such features as Locative and Instrument can be attached to nodes whose labels are more general than the features, but which capture what the corresponding observable facts have in common.

Among the advantages of such a system I would mention the following. It would serve to explain why, in various situations, there are no clear reasons to select a Locative preposition such as 'in' rather than an Instrumental preposition such as 'by' or 'with' in noun-phrases with 'machine', 'system', and the like. In addition, the proposal just outlined might serve to account for the use of the clearly Locative preposition 'on' with clearly Instrumental verbs such as 'be based on', 'depend on', and 'rely on'.

3.5.6. Anaphoric and cataphoric uses of 'system'.

I wish to end my analyses with a discussion of another remarkable property of NP's with words like '<u>system</u>' , namely that of referring to some (let us say provisionally) 'state of affairs' mentioned before or after the occurrence of the word in question, either in the same sentence or in the same text. Note, incidentally, that various other NP's have the same property : for instance, "on these bases" in "<u>On these bases</u>, the paper presents a new system..." (J 963). What is particularly remarkable is the anaphoric use of "on this basis", "this system...", "this technique" WITHOUT any previous occurrence of the same word in a given text. (Cp. : "Current status and needs <u>in the field</u> as viewed by practising biologists", J 220). In order to try to understand the conditions under which NP's with '<u>system</u>' are used to refer forward or backward to something in the same discourse, let us first consider imaginary examples of anaphoric (65) and cataphoric (66) uses of this word. Let A and B stand for two different persons.

(65)

  (a) A : You know, John broke the window with his gun.
      B : That's a good system.
  (b) A : You know, John used his gun to break the window.
      B : That's a good system.
  (c) A : You know, our machine does all the work.
      B : That's a good system.
  (d) x A : You know, it rains a lot in Britain.
      B : That's a good system.
  (e) x A : You know, three months elapsed.
      B : That's a good system.
  (f) x A : You know, John broke the window.
      B : That's a good system.
  (g) A : You know, John broke the window to get in.
      B : That's a good system.
  (h) x A : You know, he presents his show to-morrow.
      B : That's a good system.
  (i) A : You know, he presents his show to-morrow, because otherwise he would appear as the main suspect.
      B : That's a good system.

(66)

Let me tell you about a new system :

(a) John broke the window with his gun.
(b) John used his gun to break the window.
(c) Our machine does all the work.
(d) x It rains a lot in Britain.
(e) x Three months elapsed.
(f) x John broke the window.
(g) John broke the window to get in.
(h) x He presents his show to-morrow.
(i) He presents his show to-morrow, because otherwise
    he would appear as the main suspect (= he presents
    his show to-morrow in order not to appear as the
    main suspect = by presenting his show to-morrow
    he will not appear as the main suspect.)

(65) and (66) suggest that both the anaphoric and the
cataphoric uses of an NP with '<u>system</u>' or the like depend on
the presence of what I propose to represent as a pair consist-
ing of a Cause-clause and a Goal (as in b, g, and i), or on
the presence of an Instrument in Fillmore's sense, as in
(a), (b), and (c). The Instrument needs to be explicitly
mentioned, as suggested by the unacceptability of (f).
On the other hand, it seems to me that the Cause-Goal relation
need not be explicit, and that (f) and (h) are acceptable
if some implicit Goal is understood. (cp. g and i).

In (67) and (68), I give attested examples of respectively
anaphoric and cataphoric uses of NP's with '<u>system</u>' or the like
in the same sentence; I am not acquainted with any syntactic
rule or rules that would account for the ways in which '<u>system</u>'
can be moved away from what it is related to in such sentences.
The main items which I assume '<u>system</u>' or the like to be
related to are capitalized :

(67)

(a) The RETRIEVAL described here is a large-scale <u>system</u> and makes use of a large number of separate <u>machines</u> (J 684).

(b) SEARCHING the film is accomplished by the use of two projectors and two photodiode networks, an <u>arrangement</u> said to permit 16 simultaneous searches in ten minutes with a collection of 125,000 references (J 244).

(68)

(a) A description of the <u>system</u> and an outline of the main classes of the Barnard CLASSIFICATION (J 59).

(b) Elements of classification are introduced by such <u>devices</u> as INSERTING AC before all types of aircraft (J 1032).

(c) The system described employs three established <u>techniques</u> : MICROFILMING, PUNCHED-CARDS, and XEROGRAPHY (J 35).

The first two words ("Describes <u>how</u>") could be omitted from the abstract quoted below in (69) without making the occurrence of '<u>this system</u>' odd or unacceptable. This suggests that anaphoric '<u>system</u>' in texts has nothing to do with 'manner' ("<u>how</u>") :

(69)

Describes how the research reports of the Chemical Department of the du Pont Company are indexed and stored on punched cards, and the indexed information is retrieved by using an IBM 101 Electronic Statistical Machine. <u>This system</u> handles large numbers of chemical compounds and polymers by their structural and compositional features rather than by their names. (J 275).

The quotations in (70) show that anaphoric '<u>system</u>' in texts can be an exponent of any case : presumably an Object of '<u>maintain</u>' in (a), an Instrument in (c), and an Instrument and a Locative or both in (b). The only difference is that the Object with '<u>system</u>' in (a) is obligatory, whereas the NP's with '<u>system</u>' in (b) and (c) could be omitted without affecting either grammaticality or meaning except in subject position (b).

(70)

(a) Scientific documentation -- the operation of bringing the facts elucidated by scientific research from those who produce them to those who can use of them -- is getting out of hand. Yet we still maintain the traditional <u>system</u>. (J 97).

(b) Describes a new experimental attempt to develop automatic classifications of chemical concepts and interrelations in the disclosures of the patent literature. <u>In this system</u>, the classification system is made from the disclosures in the patent literature, thereby making it possible to search for all of the information in the file of interest to the searcher. The <u>system</u> uses three types of punched cards. (J 19).

(c) Using the officially accepted International System of Notation of Organic Compounds of the International Union of Pure and Applied Chemistry, a modified IBM 026 Card Punch with document writing feature and an IBM 866 electric typewriter, it is now possible to assign to each organic compound a single, mathematically unique group of descriptors, and to punch these into a card and retrieve the original cipher in typewritten form at the rate of 600 characters per minute. <u>By means of this system</u>, it is also possible to search i r any fragment or combination of fragments, either actually or potentially present in the structure. (J 273).

Also about the texts just quoted, let us note that they confirm the hypothesis, according to which anaphoric 'system' refers to what I analyze as SO--> SCause + SGoal : for instance, '<u>attempt</u> (made) (= Cause-S) <u>to develop</u> ...'(= Goal-S). In fact, the following examples suggest that one of the important functions of deictic NP's with 'system' in discourse i3 precisely that of linking together Cause and Goal sentences or clauses :

(71)

(a) A letter to the editor proposing the following <u>system</u> of communicating research results : The contributor prepares a full account of his research, sparing no detail, which is permanently filed in a central repository... [and] a two-page abridgment, which is published in journal form and numbered for identification, and a conventional abstract, which is reproduced on cards (border-punched preferably). A subscriber may receive the journal and the cards or, if he prefers, just the cards and may order by postcard photocopies or facsimile reproductions of the full papers on deposit. When the papers prove to be exceptionable, the recipients should voluntarily send their suggestions to the authors, who may then issue revisions to supersede their earlier accounts. An editor may select for full publication articles worthy of general attention or those for which the demand exceeds the resources of economic facsimile copying. (J 783).

(a') A letter to the editor proposing a system FOR communicating research results WHEREBY (IN WHICH) the contributor... <u>and</u> WHEREBY (IN WHICH) a subscriber ... <u>and</u> WHEREBY (IN WHICH) the recipient... <u>and</u> WHEREBY (IN WHICH) an editor...

(b) It should be possible to set up, in the neat future, a <u>procedure whereby</u> one can either subscribe to scientific journals on a yearly basis or purchase single articles at a unit price. <u>Such a system</u> would give scientists greater opportunity to select reading material on the basis of subject and would provide greater return for publishers. It would also make the new scanning tools more available by increasing the availability of items discovered in them. (J 629).

(b') It should be possible to set up... a procedure WHEREBY one can either subscribe... IN ORDER TO give scientists greater opportunity to select reading material and to make the new scanning tools more available...

In the texts quoted in (71), we can attempt to convert
all the constructions presumably relatable to relative
clauses . (such as "system of communicating") and relative
clauses (such as "system whereby/in which ...") into 'inde-
pendent' sentences with "system". This can be done on the
basis of suggestions put forward at the end of 3.5.4., and
of the Locative status of the word "system". Such paraphrases
are given in (72), together with clearly unacceptable
or dubious paraphrases :

(72)

    (a) A letter to the editor proposing a (new ?)
        system. This system would be USED to communicate
        research results. The system would be BASED
        on the preparation by the contributor..., on the
        possibility for the subscriber to receive ....,
        on the obligation (?) for the recipient to
        send..., and on the possibility for an editor
        to select...

    (a') A letter to the editor proposing a (new ?)
        system. This system would be USED to communicate
        research results. IN this system...

    (a") ? A letter ... proposing a (new ?) system.
        This system would COMMUNICATE research results.
        BY (MEANS OF) this system, the contributor
        prepares...

    (a"') xThis system prepares a full account...

    (b) It should be possible to set up, in the near
        future, a (new ?) procedure. The procedure would
        be BASED on the possibility of subscribing...
        Such a system would (SERVE TO) GIVE scientists
        greater opportunity to ...

    (b') It should be possible to set up, in the near
        future, a (new ?) procedure. IN this system
        (procedure ?), one can subscribe...

    (b") ? BY MEANS OF this system one can subscribe...

    (b"') xThis system subscribes...

These paraphrases suggest that :

(i) the constructions with 'BASED on' are acceptable;

(ii) the Locative use of '<u>system</u>' ('IN this system) makes it possible to use co-ordinate sentences instead of complex sentences with 'BASED on';

(iii) '<u>IN this system</u>' can be used when (presumably) Instrumental uses of '<u>system</u>' are dubious ('By MEANS OF this system') or ruled out ('x This system subscribes...');

(iv) <u>WHEREBY</u> (cp. IN WHICH) clauses correspond to 'independent' sentences with 'IN this system'.

On the basis of what precedes, I would conclude the following. What I propose to analyze as Cause-S can be realized in discourse as relative clauses with WHEREBY or IN WHICH depending on '<u>system</u>', as one 'independent' clause with 'BASED on' or the like as main verb and '<u>system</u>' as subject, or as any number of co-ordinate sentences with an <u>optional locative prepositional</u> phrase such as '<u>In this system</u>' acting as discourse connective. I am unable to represent satisfactorily the Locative nature of 'IN WHICH' and 'IN this system' on the basis of my proposals about asymmetric conjunction. What I propose to analyze as Goal-S can be realized as a nominal construction ('system of/for communicating'), or as an independent sentence with USE or the like ('serve to', etc.) or an equivalent 'simple' sentence with '<u>system</u>' as Instrument in Fillmore's sense when the verb tolerates such an Instrument.

Lastly, I wish to mention a difficulty already referred to about surface conjunctions which also arises in texts. The difficulty is that of determining the semantic status of sentences relatable to (cataphoric) '<u>system</u>'(or the like) because the text contains no mark (such as prepositions) for deciding that the sentences in question are to be analyzed as Cause or Goal. Thus, in the passage quoted below in (73), there is an obvious cataphoric relation between

"three ... techniques", and "microfilming, punched-cards, and xerography" on the one hand, and the sentences that follow on the other. It is also clear from these sentences that time sequence is involved in the enumeration "microfilming, punched-cards, and xerography", and in the corresponding sentences that follow. But ̄ am unable to identify clues that would make it possible to analyze the text in terms of the Cause-Goal relation proposed in this thesis. The word and sentence order, and the "then" in the last sentence do not provide clear enough criteria to set up a detailed AND motivated analysis of the passage quoted below:

(73)

The system described employs three established techn ̣ ̣ues : microfilming, punched-cards, and xerography. Documents to be held on file are microfilmed. Each individual picture is cut off ̣ ̣e reel of microfilm and mounted over a special aper-ture cut in a standard eighty-column punched-card. Data identifying the microfilm are punched in columns available on the card. When the docu-ment is to be reproduced, the appropriate card is selected either manually or mechanically from the file of cards and placed in a special projector mounted on a xerographic printer. The required number of full-size documents is then printed auto-matically at a speed of twenty feet a minute. (J35).

Note, however, that the second sentence ("Documents to be held on file are microfilmed") could be introduced by 'In this system'. Note furthermore that (73) could be paraphrased as in (74) :

(74)

Describes a system which employs three techniques (...) (and) WHEREBY (IN WHICH) documents to be held on file are microfilmed, WHEREBY (IN WHICH) each individual picture is cut off....

APPENDIX 1.

**MAIN HEADINGS AND DEFINITIONS OF A LIST OF DESCRIPTORS
FOR THE FIELD OF SCIENTIFIC AND TECHNICAL INFORMATION (x).**

Groupe 1 : I'INFORMATION SCIENTIFIQUE ET TECHNIQUE :
  GENERALITES.

Définitions de l'IST ; son objet, ses buts, sa place dans
l'activité nationale, dans l'économie, etc... Ses rapports
avec la science et l'industrie.
Histoire de l'IST ; son développement récent, la "crise"
actuelle de l'IST ; considérations sur l'avenir de l'IST.

Groupe 2 : ORGANISATION DE LA PROFESSION.

Considérations générales sur la ou les professions opérant
dans le domaine de l'IST (bibliothécaires, documentalistes,
traducteurs, analystes, etc...) ; rôle, développement,
réformes souhaitables, etc...
Description de l'organisation professionnelle dans tel ou
tel cadre concret : formation, recrutement, enseignements
spécialisés, niveaux académiques, salaires, etc... ;
associations professionnelles (nationales, internationales).

Groupe 3 : LA SCIENCE DE L'IST.

Etudes théoriques sur les méthodes qui font de l'IST un
domaine scientifique ou technique particulier ; l'IST
considérée comme une science (la "science de l'information");
les matières qui la constituent.
Doctrines, théories, écoles de pensée, sur la manière de
formuler ou de résoudre les problèmes intellectuels de
l'IST (notamment ceux qui font l'objet des groupes 12 à
19) ; les modèles correspondants.

Groupe 4 : SCIENCES ANNEXES.

Les sciences auxquelles l'IST fait appel (mathématiques,
cybernétique, sémiologie, etc.) ; les concepts ou méthodes
qui leur sont propres, considérés dans leur application à
l'IST.

(x) All the footnotes have been left out in the French
definitions given in this appendix, which are quoted
from N. Gardin and F. Lévy (1967).

Groupe 5 : ORGANISATION GENERALE DE L'IST.

Ensemble des organismes et institutions qui forment l'appareil général de l'IST, dans un ou plusieurs pays, ou un plusieurs domaines d'application ; sous-ensembles particuliers, différemment caractérisés (par leur type d'activité, par leur domaine linguistique, par leur statut juridique, etc.); centres nationaux ou internationaux considérés dans leurs rapports avec les organismes dont ils dirigent ou coordonnent les activités ; coopération inter-centres, ses modalités.

Dispositions politiques et juridiques déterminant la structure et le fonctionnement des réseaux d'information.

Groupe 6 : ORGANISMES D'IST.

Organismes à vocation plus ou moins large dans le domaine de l'IST (archives, services d'analyse, centres de traduction, etc.), étudiés individuellement - structure, fonctionnement, coûts, budgets, etc. - ou par types : rôles, modèles, évolution, notamment en rapport avec la mécanisation (ex. : "la bibliothèque en l'an 2000", etc.).

Groupe 7 : FONCTIONNEMENT DE L'IST.

Circulation de l'IST, des "producteurs" (auteurs, éditeurs) aux "consommateurs" (usagers) en passant par les "transformateurs" (analystes, etc.) ; appréciations sur le fonctionnement des réseaux d'information considérés au groupe 5 (i.e. leur valeur dynamique, et non plus seulement leur description statique).

Comportements, habitudes, besoins, désirs des usagers quant aux formes et quant à l'utilisation de l'IST.

Etudes économiques sur le coût de l'IST, considérée comme un service public ; sa part dans le budget national.

Groupe 8 : SOURCES D'INFORMATION.

Les différentes catégories de documents primaires, selon leur forme matérielle (textes, cartes, objets, etc.), les modalités de leurs publications (articles, actes de congrès, mélanges, etc.), etc. : inventaires, évaluations, coûts, dispositions légales, etc...

Les différentes catégories de documents secondaires (ou tertiaires, etc.) - index, bibliographies, résumés, exposés de synthèse, etc. - indépendemment des organismes chargés de les produire (groupe 6) et des méthodes suivies pour les élaborer (groupes 12 à 17) : inventaires, évaluations, coûts ; dispositions légales, etc.

Groupe 9 : POLYGRAPHIE ET REPROGRAPHIE.

Procédés et catégories d'équipements utilisés pour la re-
production de documents (impression, copie), y compris la
micro-reprographie (miniaturisation, agrandissements),
indépendemment des organismes qui les utilisent (cf.
groupe 6), ou des évaluations d'usagers dont ils font
l'object (cf. groupe 7) ; leurs coûts.
Problèmes juridiques posés par l'emploi de ces techniques.

Groupe 10 : TRAITEMENT DE L'INFORMATION : GENERALITES.

Ensemble des opérations dont les documents ou informations
scientifiques font l'objet, depuis leur collecte (groupe
11) jusqu'à leur dissémination (groupe 22), en passant
par les phases intermédiaires de l'analyse (groupes 12 à
19) et de la recherche documentaires (groupe 21) : vues
synthétiques sur la chaîne de traitement, modèles gé-
néraux, systèmes globaux.
La mécanisation des chaînes de traitement, lorsque les
quatre phases ci-dessus sont visées : collecte, analyse
(sous quelque variante que ce soit : cf. groupes 12 à 19),
exploitation et dissémination ; sinon, cf. les groupes
correspondants à chacune de ces phases.

Groupe 11 : COLLECTE DES INFORMATIONS.

L'acquisition des documents ou des informations, dans les
services de stockage (bibliothèques, archives) ou d'analyse
(centres de documentation) : problèmes, méthodes, procédés
de gestion (mécaniques ou autres).

Groupe 12 : ANALYSE DOCUMENTAIRE : GENERALITES.

Ensemble des transformations que peuvent subir les documents
ou informations scientifiques, du point de vue de leur
signalement (groupe 13) et/ou de leur contenu (groupe 14
à 19), entre le moment de leur collecte (groupe 11) et
celui de leur mise en mémoire, sous quelque forme que ce
soit (groupe 19) : vues générales sur plusieurs de ces
transformations, études synthétiques sur la chaîne
d'analyse.
Mécanisation de cette chaîne, lorsque plusieurs types
d'analyse sont considérés; sinon, cf. les groupes corres-
pondants à chacun de ces types.

Groupe 13 : CATALOGAGE.

Règles ou pratiques concernant la rédaction des titres ou
le signalement des documents pris en compte dans un
service d'IST.

Groupe 14 : CONDENSATION : RESUMES.

Au sens large, toute paraphrase d'un texte scientifique
sous forme d'un texte plus court, rédigé dans une langue
naturelle quelconque (résumés indicatifs ou informatifs,
comptes-rendus critiques ou neutres, etc.) : règles,
pratiques, méthodes, mécanisation éventuelle.

Groupe 15 : SYNTHESE(S).

Rédaction d'un document secondaire se rapportant à une
question ou à un domaine scientifique donné, à partir de
plusieurs documents primaires (exposés de synthèse,
"states-of-the-art", "trend reports", etc.) : principes
et méthodes.

Groupe 16 : CLASSIFICATION ET INDEXATION.

Caractérisation ou représentation d'un document (forme et/
ou contenu) au moyen de termes (mots-vedette, descripteurs,
termes d'indexation, etc.) tirés d'un langage documentaire
quelconque : classification, liste alphabétique de descrip-
teurs, thesaurus, etc. ; principes et méthodes.
Etudes sur des langages documentaires particuliers.
Mécanisation de l'indexation, i.e. de la conversion du
langage naturel au langage documentaire, dans quelque
domaine et pour quelque langue que ce soit.
Mécanisation de la classification, i.e. de la genèse des
langages documentaires utilisés pour l'indexation.

Groupe 17 : EXTRACTION, TABULATION : INDEX.

Observation des occurrences de certains mots, groupes de
mots ou phrases, dans les documents (langage naturel),
ou de descripteurs dans les représentations indexées de
ces documents (langage documentaire), en vue de l'établis-
sement d'extraits, d'index (tables d'occurrences, par
ordre alphabétique), ou de concordances (id., avec indica-
tion du contexte, plus ou moins étendu) ; méthodes d'obser-
vation : 1) par consultation de tables "positives" (listes
des mots ou groupes de mots à retenir comme constituants
d'un extrait ou comme entrées d'un index) ; 2) par consul-
tation de tables "négatives" (listes des mots ou groupes
de mots à exclure des extraits ou index) ; 3) par des
critères statistiques (fréquences, co-occurrences, etc.).
Observation de chaînes de citation, rétrospectives ou
prospectives : les "index de citation".
La mécanisation de ces processus.

Groupe 18 : TRADUCTION.

Conversion d'une langue naturelle en une autre (et non en un langage documentaire ; cf. groupe 16, ni la conversion ue l'une ou de l'autre en un système de notation symbolique, cf. groupe 19) : principes et méthodes, dans le contexte de l'IST.
La traduction automatique, appliquée à l'IST.

Groupe 19 : CODIFICATION.

Conversion de termes du langage naturel ou document ire vers un système de notation quelconque, pour faciliter l'enregistrement et l'exploitation des informations représentées par ces termes ; la conception de tels systèmes.

Groupe 20 : ENREGISTREMENT ET STOCKAGE.

Procédures et équipements spéciaux (par catégorie) utilisés pour la mise en mémoire des documents et/ou des produits de l'analyse dont ils ont fait l'objet (groupes 12 à 19) ; méthodes dites "conventionnelles", méthodes mécaniques.

Groupe 21 : RECHERCHE DOCUMENTAIRE.

Procédures de la recherche rétrospective (tris, compilations, sélections, etc.), envisagées soit dans l'abstrait indépendamment de. moyens matériels mis en oeuvre, soit dans le cadre d'une organisation concrète disposant de moyens "conventionnnels" et/ou "mécaniques" : modèles, systèmes de recherche rétrospective, stratégie de recherche, évaluation des résultats, études comparées (ex. : systèmes conventionnels vs. mécaniques), etc.

Groupe 22 : DISSEMINATION.

Procédures visant à une dissémination systématique des produits obtenus à la sortie d'un système de traitement de l'IST : abonnements, profils d'intérêt, dissémination sélective, etc.

Groupe 23 : MATERIEL.

Equipements immobiliers ou mobiliers utilisés pour les organismes d'IST : bâtiments particuliers, leurs aménagements ; le matériel spécial dont ils sont dotés, pour le stockage et la consultation des documents.

Ensembles mécaniques à fonctions multiples utilisés pour
le traitement de l'IST (entrée, stockage, sélection,
impression), par catégories (ex. : sélecteurs manuels,
équipements mécanographiques, sélecteurs photoélectriques,
calculateurs électroniques, etc.) ; codes et langages
symboliques correspondants ; composants particuliers.
Matériel de bureau affecté à des fonctions spéciales, dans
le traitement de l'IST - reprographie et polygraphie
mises à part : cf. groupe 9.

Groupe 24 : TERMINOLOGIE.

Travaux de lexicographie entrepris dans une ou plusieurs
langues naturelles, et pour quelque domaine que ce soit,
à des fins documentaires ou non : glossaires, lexiques,
dictionnaires, thesaurus, listes de "mots vides" (i.e.
les dictionnaires d'exclusion utilisés pour la fabrica-
tion de certains index KWIC ou autres), etc.

Groupe 25 : STANDARDISATION.

Recommandations, projets, accords visant à normaliser tel
ou tel élément, démarche, instrument, etc., dans les
travaux concernant l'IST (ex. : alphabets, codes, abrévia-
tions, etc.). Les objets visés par la standardisation
peuvent être désignés par des descripteurs tirés des
groupes appropriés (ex. : groupes 8, sources ; 14, résumés ;
16, langages documentaires, etc.), ou récapitulés directe-
ment dans ce groupe ; un troisième parti également acceptable
aurait été de renoncer à ce groupe, et de citer la notion
de standardisation à propos de chacun des objets auxquels
on l'applique.

APPENDIX 2.

ENGLISH TRANSLATION OF THE MAIN HEADINGS OF
THE GARDIN-LEVY CONCORDANCE.

1. Scientific and technical information : generalities.
2. Professional problems.
3. Information science.
4. Related sciences.
5. General organization of scientific and technical information.
6. Scientific and technical information institutions.
7. Functioning of scientific and technical information networks.
8. Sources of information.
9. Techniques of reproduction.
10. Information handling : generalities.
11. Collection of information.
12. Documentary analysis : generalities.
13. Cataloguing.
14. Abstracting.
15. Synthesis.
16. Classification and indexing.
17. Extracting, tabulation and index production.
18. Translating.
19. Coding.
20. Storing.
21. Searching.
22. Dissemination.
23. Machines and equipment.
24. Terminology.
25. Standards.

APPENDIX 3.

THE FIRST 50 ABSTRACTS OF JANASKE 1962 : EDITED VERSION.

1. THE JAPANx INFORMATIONx CENTERx OF SCIENCEx AND TECHNOLOGYx AND ITS ACTIVITIES . A GENERAL DESCRIPTION OF THE JAPANx INFORMATIONx CENTERx OF SCIENCEx AND TECHNOLOGYx WHICH WAS ESTABLISHED FOR THE PURPOSE OF CONTRIBUTING TO THE DEVELOPMENT OF SCIENCE AND TECHNOLOGY BY COLLECTING , PROCESSING , STORING , AND RETRIEVING SCIENTIFIC INFORMATION . AT THE TIME OF THIS REPORT , THE CENTER WAS PLANNING TO CHANGE OVER TO A MECHANIZED STORAGE AND RETRIEVAL SYSTEM SIMILAR TO THAT USED BY THE U.S.x PATENTx OFFICEx . SERVICES AVAILABLE FROM THE CENTER AND THE CHARGES FOR EACH ARE LISTED .

2. REPORT ON A STUDY OF SCIENTIFIC COMMUNICATION FOR THE NATIONALx SCIENCEx FOUNDATIONx . THIS REVIEWS THE INITIAL REASONING AND THINKING BEHIND THE DESIGN AND CONSTRUCTION OF A PROJECT SUPPORTED BY THE NATIONALx SCIENCEx FOUNDATIONx , CARRIED OUT BY THE OPERATIONSx RESEARCHx GROUPx AT THE CASEx INSTITUTEx OF TECHNOLOGYx TO DETERMINE WAYS TO AFFECT THE BEHAVIOR OF SCIENTIFIC INSTITUTIONS THAT ARE RESPONSIBLE FOR THE DISSEMINATION OF RECORDED INFORMATION IN ORDER TO ACHIEVE A NET IMPROVEMENT IN SCIENTIFIC PRODUCTIVITY .

3. PROBLEMS IN COMMUNICATING RUSSIANx SCIENCE . AN OUTLINE OF THE NATIONALx INSTITUTEx OF HEALTHx RUSSIANx SCIENTIFIC TRANSLATION PROGRAM , WITH A DISCUSSION OF THE PROBLEMS OF SELECTION , FORM , DISTRIBUTION , ACCEPTABILITY , AND TRANSLATION RIGHTS .

4. SEARCHING MEDICINAL CHEMICAL LITERATURE . xTHREE REPRESENTATIVE PROBLEMS OF A MANUFACTURER OF FINE CHEMICALS IN SEARCHING FOR INFORMATION IN THE MEDICAL AND CHEMICAL LITERATURE : SEARCHING THE BOTANICAL DRUG LITERATURE FOR FOLK REMEDIES ; INTERPRETING AND DETERMINING THE CONTENT OF FOREIGN PRESCRIPTIONS ; AND TRACING AND IDENTIFYING THE MERCKx REFERENCES .

5. THE FEDERALx GOVERNMENTx AND U.S.x SCIENTIFIC INFORMATION . IN ADDITION TO PRESENTING A GENERAL REVIEW OF THE SCOPE OF THE FEDERALx GOVERNMENTx 'S ROLE IN RESEARCH AND DEVELOPMENT , AND A REVIEW OF THE CHARACTER OF THE FEDERALx GOVERNMENTx 'S SCIENTIFIC INFORMATION ACTIVITIES , THIS ARTICLE PRESENTS A REVIEW OF THE xSIX-POINT PROGRAM WHICH HAS BEEN DEVELOPED BY THE NATIONALx SCIENCEx FOUNDATIONx FOR IMPROVING THE DISSEMINATION OF SCIENTIFIC INFORMATION , PARTICULARLY AMONG U.S.x SCIENTISTS AND ENGINEERS . THE AUTHOR CONCLUDES THAT THE DISSEMINATION OF SCIENTIFIC INFORMATION MUST BE RECOGNIZED AS AN INTEGRAL PART OF RESEARCH AND DEVELOPMENT .

6. UNITEDx STATESx SCIENTIFIC AND TECHNICAL INFORMATION SERVICES .
THIS DISCUSSION OF UNITEDx STATESx SCIENTIFIC AND TECHNICAL
INFORMATION SERVICES IS DIVIDED INTO THE FOLLOWING PARTS :
THE PRESENT SITUATION IN THE UNITEDx STATESx ; SOME OF THE
REASONS BEHIND THE CURRENT PROBLEMS ; THE PRINCIPAL SCIENTIFIC
AND TECHNICAL INFORMATION ACTIVITIES OF THE FEDERALx
GOVERNMENTx ; AND A PROPOSED PROGRAM TO REMEDY THIS COUNTRY
'S SCIENTIFIC AND TECHNICAL INFORMATION ILLS .

7. LIST OF DOCUMENTATION CENTERS IN THE GERMANx DEMOCRATICx
REPUBLICx . THE ADDRESS IS GIVEN FOR EACH OF x82 DOCUMENTATION
CENTERS . ADDITIONAL INFORMATION CONCERNS AVAILABILITY OF
MICROFILM AND PHOTOCOPYING SERVICES OR PHOTOCOPYING SERVICE
ONLY , THE SUBJECT OF PRIMARY INTEREST TO THE CENTER , AND
COVERAGE PROVIDED OVER PERTINENT SOVIETx PERIODICAL LITERATURE.

8. TRAINING THE SCIENTIFIC INFORMATION OFFICER . POINTS OUT THE
DIFFERENCE BETWEEN THE FUNCTION OF A SCIENTIFIC INFORMATION
OFFICER AND A LIBRARIAN . A SYLLABUS IS PRESENTED FOR A
POST-GRADUATE COURSE OF TRAINING FOR A STUDENT WHO IS ALREADY
A SUBJECT SPECIALIST .

9. THE IBM-650x INFORMATION RETRIEVAL SYSTEM . DESCRIBES THE
IBM-650x INFORMATION RETRIEVAL SYSTEM WHICH USES THE INVERTED
FILE . IN THIS FILE EACH TERM RECORD CONTAINS ALL OF THE
DOCUMENT CODES FOR ALL OF THE DOCUMENTS IN WHICH THE TERM
APPEARS . THEREBY IT IS ONLY NECESSARY FOR THE MACHINE TO
SEARCH THE PERTINENT TERMS IN THE MACHINE 'S MEMORY.

10. STORAGE AND RETRIEVAL OF BIOLOGICAL INFORMATION . A GENERAL
DISCUSSION OF THE PRESENT STATE OF BIOLOGICAL LITERATURE ,
WITH EMPHASIS ON SCIENTIFIC NOMENCLATURE , THE LIFE SPAN OF
BIOLOGICAL LITERATURE , AND THE STORAGE AND RETRIEVAL OF
BIOLOGICAL SPECIMENS .

11. SOME PROBLEMS AND QUESTIONS OF PATENT DOCUMENTATION IN INDUSTRY .
A COMPARISON IS MADE AMONG MAGNETIC DRUMS , MAGNETIC TAPE ,
MAGNETIC WIRE , FERROMAGNETIC CORES , PUNCHED-CARDS , PUNCHED
TAPES , AND PHOTOGRAPHIC TAPES WITH RESPECT TO THEIR ACCESS ,
DURABILITY , STORAGE AND RETRIEVAL TIME , CAPACITY , AND
EFFICIENCY . IT IS CONCLUDED THAT THE MOST ECONOMIC , AS FAR
AS COSTS PER INFORMATION UNIT ARE CONCERNED , IS MAGNETIC
TAPE . IT ALSO POSSESSES THE GREATEST INFORMATION CAPACITY .
xFIVE SUGGESTIONS ARE ADVANCED FOR A PRACTICAL SOLUTION
OF PATENT DOCUMENTATION PROBLEMS .

12. TOWARD A PROCEDURE FOR LOGICALLY CATALOGING KNOWLEDGE .
SUGGESTS AN APPROACH TO THE CATALOGING OF THE ENTIRE BODY OF
WRITTEN KNOWLEDGE BY VIEWING IT AS FINITE SERIES OF CONJOINED
PROPOSITIONS WHICH MAY BE ORDERED ACCORDING TO THEIR LOGICAL
STRUCTURE . AN APPLICATION OF THIS METHOD TO CATALOGING THEOREMS
OF MODERN LOGIC IS SELECTED AS PARTICULARLY APPROPRIATE FOR
ILLUSTRATIVE PURPOSES . BINARY-OCTAL DESIGNATIONS ARE
ESTABLISHED FOR THEOREMS , SYSTEMS OF LOGIC AND FOR STATEMENTS
IN PROPOSITIONAL CALCULUS , DERIVED FROM CERTAIN OF THEIR BASIC
CHARACTERISTICS . THE AUTHOR RECOGNIZES THAT NOT ALL STATEMENTS
CAN BE EXPRESSED IN AS NEAT A FASHION AS HIS EXAMPLES FROM
THE FIELD OF LOGIC . HE FEELS , HOWEVER , THAT AT LEAST
SUBSTANTIAL PROPORTION OF THAT WHICH MAY BE WRITTEN IN A
NATURAL LANGUAGE CAN BE REDUCED TO A STANDARD FORM THAT
WILL DETERMINE THE APPROPRIATE PLACEMENT IN A CATALOG .

13. SEARCHING THE CHEMICAL LITERATURE . THIS IS A REVISED AND
ENLARGED EDITION OF ADVANCESx IN CHEMISTRYx SERIESx . BASED
ON PAPERS PRESENTED BY THE DIVISIONx OF CHEMICALx LITERATUREx
AND THE DIVISIONx OF CHEMICALx EDUCATIONx OF THE AMERICANx
CHEMICALx SOCIETYx AT NATIONAL MEETINGS FROM x1947 TO x1956 .
CONCERNED PRIMARILY WITH CHEMICAL INFORMATION STORAGE AND
RETRIEVAL , MOST OF THE MATERIAL HAS BROAD APPLICATIONS IN
OTHER DISCIPLINES .

14. ROSTER OF CURRENT RESEARCH , DEVELOPMENT , AND TESTING IN
DOCUMENTATION AND LIBRARIANSHIP , x1957-1958 . LISTS APPROXIMA-
TELY x80 INDIVIDUALS OR ORGANIZATIONS WORKING IN THE FOLLOWING
x7 AREAS : CODING FOR MECHANICAL SEARCHING SYSTEMS ; RETRIEVAL
AND REPRODUCTION ; EQUIPMENT FOR INFORMATION STORAGE ; INDEXING
, CATALOGING , AND CLASSIFICATION ; TRANSLATION BY MECHANICAL
MEANS ; PRODUCTION AND DISSEMINATION OF INFORMATION ; USE OF
INFORMATION AND USER REQUIREMENTS ; THEORETICAL STUDIES .

15. ADVANCES IN EDPx AND INFORMATION SYSTEMS . PAPERS IN THIS
VOLUME ARE BASED ON MATERIAL ORIGINALLY PRESENTED AT AMASx
'S xSEVENTH ANNUALx DATAx PROCESSINGx CONFERENCEx , HELD
IN MARCHx x1961 . PARTIAL CONTENTS : ADVANCES IN INFORMATION
RETRIEVAL AND DATA ACQUISITION ; PROGRESS IN THE DESIGN OF
INFORMATION RETRIEVAL SYSTEMS ; IMPROVED INFORMATION STORAGE
AND RETRIEVAL SYSTEMS .

16. TECHNICAL COMMUNICATION IN PSYCHOLOGY ; A STATEMENT OF THE
PROBLEM . THIS STATEMENT EVOLVED OUT OF THE DELIBERATIONS
OF THE BOARDx OF SCIENTIFICx AFFAIRSx OVER A SPAN OF xTWO
YEARS . THE BSAx CONSIDERS THE PROBLEM OF EFFICIENT AND
EFFECTIVE COMMUNICATION OF SCIENTIFIC INFORMATION TO BE PERHAPS
THE MOST CRITICAL PROBLEM FACED BY SCIENTIFIC PSYCHOLOGY TODAY .
THE PRINCIPAL NEEDS OF SCIENTIFIC PSYCHOLOGY FOR COMMUNICATION
MAY BE CATEGORIZED INTO xFOUR CLASSES : NEED FOR RAPID COM-
MUNICATION OF WHAT AND HOW , NEED FOR DIRECT DISCUSSION AND
COMPARISON OF IDEAS AND FINDINGS , NEED FOR ADEQUATE AND
EFFICIENT ARCHIVAL STORAGE , AND NEED FOR INTEGRATED RETRIEVAL
OF INFORMATION . WITH RESPECT TO THE LATTER PSYCHOLOGISTS
SHOULD ATTEND IMMEDIATELY TO THE PROBLEM OF ENCYCLOPEDIC

ORGANIZATION OF THEIR KNOWLEDGE AND CODIFICATION OF THE METHODS, MEASURES AND RESULTS OF PSYCHOLOGY SO THAT ADVANTAGES OF THE NEW TECHNIQUES FOR AUTOMATIC STORAGE OR RETRIEVAL OF INFORMATION MAY BE EXPLOITED .

17. RESEARCH STUDY OF CRITERIA AND PROCEDURES FOR EVALUATING SCIENTIFIC INFORMATION RETRIEVAL SYSTEMS . A STUDY OF THE REQUIREMENTS , CRITERIA , AND MEASURES OF PERFORMANCE WHICH MAY BE USED TO EVALUATE AND COMPARE SCIENTIFIC INFORMATION RETRIEVAL SYSTEMS . BASED ON THE RESULTS OF THIS STUDY , RECOMMENDATIONS WERE MADE FOR APPLICATION OF THESE FINDINGS TO EXISTING AND PROPOSED SYSTEMS , AND FOR FURTHER RESEARCH EFFORTS IN THE EVALUATION OF SCIENTIFIC INFORMATION RETRIEVAL SYSTEMS .

18. APPLICATION OF HIGH-SPEED COMPUTERS TO INFORMATION RETRIEVAL . THIS IS A DESCRIPTION OF THE FOLLOWING TYPES OF INFORMATION RETRIEVAL SYSTEMS : STATISTICAL , SYNTACTICAL AND INTERRELATIONAL . IN ADDITION , THE COMPUTER RESEARCH PROGRAM OF THE PATENTx OFFICEx IS DESCRIBED .

19. AUTOMATIC CATEGORIZATION OF CHEMICAL CONCEPTS AND INTERRELATIONSHIPS . DESCRIBES A NEW EXPERIMENTAL ATTEMPT TO DEVELOP AUTOMATIC CLASSIFICATIONS OF CHEMICAL CONCEPTS AND INTERRELATIONS IN THE DISCLOSURES OF THE PATENT LITERATURE . IN THIS SYSTEM , THE CLASSIFICATION SYSTEM IS MADE FROM THE DISCLOSURES IN THE PATENT LITERATURE , THEREBY MAKING IT POSSIBLE TO SEARCH FOR ALL OF THE INFORMATION IN THE FILE OF INTEREST TO THE SEARCHER . THE SYSTEM USES xTHREE TYPES OF PUNCHED-CARDS . ONE OF ITS FEATURES IS THAT ALTHOUGH THE CLASSIFICATION IS VARIABLE , THE DISCLOSURES ARE FIXED .

20. INTERRELATIONSHIPS AS A BASIS FOR INFORMATION RETRIEVAL . THE WRITER FEELS THAT NO LARGE SCALE INFORMATION RETRIEVAL SYSTEM CAN BE A SUCCESS UNLFSS IT CAN HANDLE SEARCHING FOR COMPLEX RELATIONSHIPS . HE EXPLAINS SOME SCHEMES FOR CODING RELATION - SHIPS USING SEPARATE RELATIONSHIP ITEMS ; COMPOUNDING ITEMS ; ORDERING ITEMS ; AND INTERFIXING TERMS .

21. PROBLEMS OF PREPARING AN INVERTED FILE . AN INVERTED FILE IS AN ORGANIZING SYSTEM FOR THE STORED INFORMATION IN A MACHINE MEMORY SO THE INFORMATION IS DIVIDED INTO AS MANY SECTIONS AS THERE ARE DIFFERENT CODES REPRESENTING ALL OF THE SUBUNITS OF INFORMATION IN ALL OF THE DOCUMENTS IN THE MEMORY OF THE MACHINE . THE MEMORY ADDRESSES USED TO DESIGNATE THESE SECTIONS IN A PARTICULAR MACHINE SYSTEM ARE USED AS THE CODES FOR THE INFORMATION SUBUNITS . THE ADVANTAGES OF THIS INVERTED SEARCH FILE AS COMPARED TO SOME OTHER SEARCH FILE SYSTEM ARE MANY , INCLUDING LESS TIME FOR MACHINE INTERROGATION OF THE FILE . MANY OF THESE ADVANTAGES WILL BE LOST UNLESS A SYSTEMATIC PROCEDURE IS DEVELOPED FOR PREPARING THE FILE . THE AUTHOR SUGGESTS WAYS OF ATTACKING THE PROBLEMS CONFRONTED IN DEVELOPING SUCH A FILE PREPARATION PROCEDURE .

531.

22. RECENT ADVANCES IN PATENTx OFFICEx SEARCHING : STEROID
    COMPOUNDS AND ILASx . A DESCRIPTION OF SEVERAL METHODS OF
    CODING CHEMICAL COMPOUNDS IN PATENTS FOR   THE PILOT PROJECT
    OF THE U.S.x PATENTx OFFICEx AND THE NATIONALx BUREAUx
    OF STANDARDSx . THE CODED INFORMATION IS SEARCHED BY MACHINE
    TO PROVIDE THE EXAMINER WITH PRIOR ART SEARCH . THE RESULTS
    ARE GRATIFYING IN THAT IT IS PROVING TO BE POSSIBLE TO
    REDUCE THE EXAMINER 'S TIME REQUIRED TO COMPLETELY HANDLE A
    PATENT APPLICATION FROM AN AVERAGE OF xTEN HOURS EACH TO xTWO
    HOURS OR LESS . IN ADDITION , IT APPEARS THAT THE RESULTS
    OF THE MACHINE SEARCHES WILL BE MORE ACCURATE THAN HAND
    SEARCHES .

23. BASIC DIRECTIONS OF WORK AT THE EXPERIMENTALx LABORATORYx
    OF MACHINEx TRANSLATIONx . THIS IS AN EXPLANATION OF THE
    PROGRAM OF RESEARCH CARRIED ON AT THE LENINGRADx UNIVERSITYx
    EXPERIMENTALx LABORATORYx OF MACHINEx TRANSLATIONx . ONE
    BASIC PART OF THEIR PROGRAM INVOLVES THE USE OF AN INTERMEDIATE
    LANGUAGE BETWEEN THE INPUT AND OUTPUT LANGUAGES .

24. THE UNIVERSAL CODE OF SCIENCE AND MACHINE LANGUAGES . THE
    CREATION OF A UNIVERSAL CODE OF SCIENCE STILL WILL REQUIRE MUCH
    WORK . xTWO OF THE ELEMENTS NEEDED ALREADY EXIST IN A PARTIAL -
    LY DEVELOPED FORM . WORK IS JUST BEGINNING ON THE DEVELOPMENT
    OF THE xTHIRD ELEMENT , MACHINE LANGUAGES .

25. ARCHIVES , LIBRARIES AND DOCUMENTATION CENTERS OF SWITZERLANDx..
    GUIDE THROUGH SWISSx DOCUMENTATION . AN ANNOTATED INDEX TO
    x459 OF THE MOST IMPORTANT SWISSx LIBRARIES , ARCHIVES ,
    DOCUMENTATION CENTERS , ETC. , ARRANGED BY GEOGRAPHICAL LOCA-
    TION . DESCRIPTIVE DATA INCLUDE HISTORICAL DEVELOPMENT ,
    HOLDINGS , SPECIAL COLLECTIONS , CLASSIFICATION SYSTEMS USED ,
    AND USER INFORMATION . THE ANNOTATIONS ARE WRITTEN IN THE
    LANGUAGE USED IN THE PERTINENT GEOGRAPHICAL REGION : GERMANx ,
    FRENCHx , ITALIANx . ALPHABETICAL INDEXES OF INSTITUTIONS ,
    SUBJECTS , AND PERSONNEL ARE INCLUDED .

26. A xTHREE-SYMBOL CODE FOR SEARCHING CHEMICAL STRUCTURES .
    A DESCRIPTION OF A xTHREE-SYMBOL CODE DESIGNED FOR USE AS A
    DETAILED SEARCH CODE FOR PHARMACEUTICAL CHEMICALS . THIS
    IS A MODIFICATION OF THE WISELOGLEx CLASSIFICATION SYSTEM .
    THE CODE AND CARD DESIGN CONSTITUTES A MAJOR CHANGE , SINCE
    ALL FUNCTIONAL UNITS , AND NOT MERELY THE FILING CATEGORIES ,
    CAN BE PUNCHED , PROVIDING A SEARCH CODE AS WELL AS A FILING
    ORDER CODE . ALTHOUGH THIS SYSTEM WAS ORIGINALLY DESIGNED
    TO BE USED WITH EDGE-PUNCHED , HAND-SORTED CARDS THE AUTHOR
    INDICATES THAT IT CAN BE EASILY MODIFIED FOR CENTER-PUNCHED
    CARD WHICH WOULD PERMIT MACHINE SORTING .

532.

27. INFORMATION RETRIEVAL : A PRAGMATIC APPROACH . A DESCRIPTION
    OF THE DEVELOPMENT OF A NEW SUBJECT INDEX FOR THE LIBRARYx
    AND INFORMATIONx SERVICEx , PAINTx DIVISIONx , IMPERIALx
    CHEMICALx INDUSTRIESx LIMITEDx . THE ARTICLE ALSO INCLUDES
    A SHORT ACCOUNT OF THE VARIOUS TYPES OF SUBJECT INDEXES
    USED BY THIS LIBRARY IN THE PAST .

28. ECONOMIC ASPECTS IN THE DISSEMINATION OF CHEMICAL KNOWLEDGE .
    PROPOSES THE ESTABLISHMENT BY THE AMERICANx CHEMICALx
    SOCIETYx OF A SYSTEM FOR HANDLING INFORMAL LITERATURE , AND
    A COMPREHENSIVE COPYING SERVICE . THIS INFORMAL LITERATURE
    WOULD BE PROVIDED IN MANUSCRIPT FORM PREPARED TO SPECIFICA-
    TIONS SUCH AS THOSE USED FOR DIVISIONAL REPRINTS AND GENERALLY
    MEETING THE STANDARDS OF COMPANY REPORTS AND MEMORANDA .

29. SYSTEMS OF SCIENTIFIC AND TECHNICAL INFORMATION SERVICES :
    LONG RANGE PLANNING . A PROPOSAL FOR THE ESTABLISHMENT OF
    AN INTERNATIONAL TECHNICAL INFORMATION SYSTEM WHICH WOULD BE
    BASED ON xTHREE MAIN INTERNATIONAL CENTERS : ONE FOR THE
    AMERICASx ; ONE FOR EUROPEx , THE NEARx EASTx AND AFRICAx ;
    ONE FOR ASIAx AND AUSTRALASIAx.. IN ADDITION , THERE WOULD
    BE A NATIONAL CENTER FOR EACH COUNTRY , AND A CENTER FOR
    EACH SCIENTIFIC OR TECHNICAL DISCIPLINE . THE PAPER ALSO
    INCLUDES DISCUSSION OF PROBLEMS WHICH WOULD BE ENCOUNTERED
    IN THE DEVELOPMENT OF SUCH A SYSTEM .

30. DIRECTORY OF DATA PROCESSING SERVICE CENTERS . LISTS SERVICE
    BUREAUS IN CANADAx AND THE UNITEDx STATESx BY PROVINCE OR
    STATE , AND BY CITY WITHIN PROVINCE OR STATE .

31. ASLIBx AEROx GROUPx CONFERENCEx . SUMMARY OF SYMPOSIUM
    ON AERONAUTICAL DOCUMENTATION IN THE SEVERAL NATOx
    COUNTRIES , INCLUDING DESCRIPTION OF EFFORTS IN CANADAx ,
    FRANCEx , HOLLANDx , FEDERAL REPUBLICx OF GERMANYx ,
    U.S.A.x , AND ITALYx .

32. THE ROLE OF LARGE MEMORIES IN SCIENTIFIC COMMUNICATION .
    LARGE MEMORIES PROVIDE AUTOMATIC REFERENCE TO xMILLIONS OF
    WORDS OF MACHINE-READABLE CODED INFORMATION OR TO xMILLIONS
    OF IMAGES OF DOCUMENT PAGES . HIGHER DENSITIES OF STORAGE
    WILL MAKE POSSIBLE LOW-COST MEMORIES OF xBILLIONS OF WORDS
    WITH ACCESS TO ANY PART IN A FEW SECONDS OR COMPLETE SEARCHES
    IN MINUTES . THESE MEMORIES WILL SERVE AS INDEXES TO THE
    DELUGE OF TECHNICAL LITERATURE WHEN THE PROBLEMS OF INPUT
    AND OF THE AUTOMATIC GENERATION OF CLASSIFICATION INFORMATION
    ARE SOLVED . DOCUMENT FILES WILL MAKE THE INDEXED LITERATURE
    READILY AVAILABLE TO THE SEARCHER . HOWEVER , MEMORY CAPACITY
    IS CURRENTLY WELL AHEAD OF OUR ABILITY TO USE IT , AND MUCH
    WORK REMAINS IN THIS AREA . MACHINE TRANSLATION OF LANGUAGES
    AND RECOGNITION OF SPOKEN INFORMATION ARE xTWO OTHER AREAS
    WHICH WILL REQUIRE FAST , LARGE MEMORIES .

3ᵀ ., SUGGESTED CLASSIFICATION FOR THE LITERATURE OF DOCUMENTATION .
AE CLASSIFICATION SCHEME PRESENTED HERE HAD ITS ORIGIN IN
A CLASS PROJECT UNDERTAKEN BY STUDENTS AT THE UNIVERSITYx
OF CHICAGOx , GRADUATEx LIBRARYx SCHOOLx DURING THE SPRING
QUARTER x1960 . A DESCRIPTION OF THE PRINCIPAL FEATURES OF
THE CLASSIFICATION WITH REFERENCE TO HAYKINx 'S xSIX BASIC
BUILDING RULES FOR SPECIAL CLASSIFICATION IS FOLLOWED BY x2
APPENDICES : CLASSIFIED GUIDE TO AMERICANx DOCUMENTATIONx ;
SUGGESTED CLASSIFICATION FOR THE LITERATURE OF DOCUMENTATION .

34. PROCEEDINGS OF THE ADIAx CONFERENCE . THIS INTERNATIONAL
MEETING WAS SPONSORED BY : THE INTERNATIONALx FEDERATIONx
FOR DOCUMENTATIONx ; THE JCINTx ENGINEERINGx COUNCILx ;
THE GERMANx DOCUMENTATIONx SOCIETYx ; AND THE GMELINx
INSTITUTEx OF INORGANICx CHEMISTRYx AND RELATEDx SCIENCESx
IN THE MAX-PLANCKx SOCIETYx FOR THE ADVANCEMENTx OF SCIENCEx .
THE PURPOSE OF THE CONFERENCE WAS TO PRESENT A CRITICAL
EVALUATION OF THE NEEDS AND POSSIBILITIES FOR EFFECTIVE
DOCUMENTATION SYSTEMS .

35. AUTOMATIC SELECTION AND REPRODUCTION OF INDUSTRIAL DOCUMENTS .
THE SYSTEM DESCRIBED EMPLOYS xTHREE ESTABLISHED TECHNIQUES :
MICROFILMING, PUNCHED-CARDS , AND XEROGRAPHY . DOCUMENTS TO
BE HELD ON FILE ARE MICROFILMED . EACH INDIVIDUAL PICTURE
IS CUT OFF THE REEL OF MICROFILM AND MOUNTED OVER A SPECIAL
APERTURE CUT IN A STANDARD xEIGHTY-COLUMN PUNCHED-CARD .
DATA IDENTIFYING THE MICROFILM ARE PUNCHED IN COLUMNS
AVAILABLE ON THE CARD WHEN THE DOCUMENT IS TO BE REPRODUCED ,
THE APPROPRIATE CARD IS SELECTED EITHER MANUALLY OR MECHANICAL-
LY FROM THE FILE OF CARDS AND PLACED IN A SPECIAL PROJECTOR
MOUNTED ON A XEROGRAPHIC PRINTER . THE REQUIRED NUMBER OF
FULL-SIZE DOCUMENTS IS THEN PRINTED AUTOMATICALLY AT A SPEED
OF xTWENTY FEET A MINUTE . THE SYSTEM IS PARTICULARLY SUITED
TO THE FILING AND REPRODUCTION OF ENGINEERING DRAWINGS
x600,000 DRAWINGS CAN BE RECORDED AND FILED IN A SPACE OF
x50 SQUARE FEET . MACHINES OF VARYING COMPLEXITY ARE AVAILABLE
FOR MOUNTING MICROFILM ON THE CARDS , THE SIMPLEST BEING
A SMALL , MANUALLY OPERATED , PORTABLE UNIT . DETAILS ARE
GIVEN OF THE COPYFLOx HIGH-SPEED PRINTER , AND THE WAY IN WHICH
THE SYSTEM IS OPERATED BY THE WESTINGHOUSEx ELECTRICx
CORPORATIONx IS DESCRIBED , WITH FIGURES OF SAVINGS ACHIEVED
IN STORAGE SPACE AND IN COSTS OF DRAWING-OFFICE SUPPLIES AND
OF REPRODUCTION DRAWINGS .

36. THE AUTOMATICx MICROFILMx INFORMATIONx SYSTEMx . A DESCRIPTION
OF THE AUTOMATICx MICROFILMx INFORMATIONx SYSTEMx AND HOW
IT WORKS . IT IS A FAST SYSTEM FOR THE RETRIEVAL OF DOCUMENTS
IN LARGE FILES WHEN THE CALL NUMBER OF THE DOCUMENT IS KNOWN .
IT IS NOT A SYSTEM FOR THE SEARCHING OF THE LITERATURE IN THE
FILES .

37. TECHNICAL INVESTIGATION OF ELEMENTS OF A MECHANIZED LIBRARY
SYSTEM . SUMMARIZES THE DESIGN AND CONSTRUCTION EFFORTS ON
FEASIBILITY MODELS OF A MEMORY WHICH STORES DOCUMENTARY INFORMA-
TION IN THE FORM OF MICROPHOTOGRAPHS ; A CAMERA DESIGNED TO
PREPARE THESE MICROPHOTOGRAPHS ; AN OUTPUT SYSTEM DESIGNED
TO DISPLAY AND REPRODUCE THE INFORMATION CONTENT OF THE MEMORY .
THE MEMORY HAS BEEN DESIGNED TO PROVIDE AN ACCESS TIME OF
BETWEEN x0.3 AND x2.0 SECONDS TO APPROXIMATELY x1,000,000
DOCUMENT PAGES ; THE CAMERA TO PRODUCE MICROPHOTOGRAPHS HAVING
LINEAR REDUCTION RATIOS OF x70 TO 1 AND x140 TO 1 , AND TO
ARRANGE THEM FOR STORAGE IN THE MEMORY   OF THE x2 OUTPUT
SYSTEMS DESIGNED , ONE STORES THE CONTENT OF A SELECTED
MICROPHOTOGRAPH TEMPORARILY IN AN ELECTROSTATIC STORAGE TUBE
AND DISPLAYS IT ON A CATHODE RAY TUBE SCREEN ; THE xSECOND
REPRODUCES ON MICROFILM THE MICROPHOTOGRAPHS SELECTED FROM THE
MEMORY . FEASIBILITY MODELS HAVE BEEN CONSTRUCTED AND ARE
UNDERGOING FINAL TESTS BEFORE BEING INSTALLED IN AN EXPERIMENTAL
SYSTEM PLANNED FOR DEMONSTRATION PURPOSES .

38. INTRODUCTION TO SCIENTIFIC DOCUMENTATION . THE BOOK COMPRISES
x4 CHAPTERS AND A SUBJECT INDEX . THE x1ST CHAPTER LAYS STRESS
ON THE ROLE AND SIGNIFICANCE OF DOCUMENTATION . THE AUTHORS
RECOUNT THE ORIGINS AND DEVELOPMENT OF DOCUMENTATION AND ITS
PRESENT ORGANIZATION ABROAD AND IN RUMANIAx , ESPECIALLY IN
SCIENCE AND TECHNICAL SUBJECTS . AN ATTEMPT IS MADE TO CLARIFY
SOME CONCEPTIONS , MAKE THE TERMINOLOGY MORE PRECISE AND TO
EXPLAIN THE DIFFICULTIES CONNECTED WITH DOCUMENTATION .
IN THE x2ND CHAPTER THE AUTHORS GIVE US THE MAIN SOURCES
OF DOCUMENTATION . THEY DISCUSS CARD-INDEXES , LIBRARIES AND
INSTITUTIONS DEALING WITH DOCUMENTATION . THE x3RD CHAPTER
POINTS OUT THE DIFFERENT KINDS OF DOCUMENTATION ACCORDING TO
THE SOURCE , CONTENT AND VALUE OF THEIR INFORMATION . THESE
KINDS ARE : GRAPHICAL , PICTORIAL , AUDITORY , VISUAL ,
ETC. . THE x4TH CHAPTER SHOWS THE STAGES OF DOCUMENTATION
INCLUDING RESEARCH WORK , PLANNING OF DOCUMENTATION , COL-
LECTING AND INVESTIGATING THE SOURCES AND USE OF DOCUMENTS .
THERE IS A VERY USEFUL SUPPLEMENT OF BIBLIOGRAPHIC TERMS IN
RUMANIANx , HUNGARIAN x , POLISHx . SERBO-CROATIANx AND
SPANISHx .

39. THE YUGOSLAVx BIBLIOGRAPHICALx INSTITUTE . DESCRIPTION
OF THE INSTITUTEx , FOUNDED IN x1948 , AND ITS SERVICES .
WORK IS CARRIED ON BY x4 DIVISIONS : BIBLIOGRAPHICALx ;
REFERENCEx AND DOCUMENTATIONx ; INTERNATIONALx EXCHANGEx OF
PUELICATIONSx CENTREx ; A'¨ TECHNICALx PRINTINGx DIVISIONx .

40. SCIENTIFIC LITERATURE USE : A SURVEY   THE AUTHOR DRAWS ,
FROM REPORTS BY BERNALx , HERNERx , AND URQUHARTx ON THE WAYS
IN WHICH SCIENTIFIC PERSONNEL GATHER INFORMATION , SOME
OBSERVATIONS AS TO CERTAIN GENERAL PRACTICES AND ATTITUDES
AMONG SCIENTISTS WHICH HE BELIEVES SHOULD BE KEPT IN MIND
BY THOSE WHO MANAGE SCIENCE LIBRARIES .

41. HOUSE ORGANS AND TRADE PUBLICATIONS AS INFORMATION SOURCES .
SUGGESTS SOLUTIONS FOR THE ACQUISITION AND HANDLING OF HOUSE
ORGANS AND OTHER TRADE PUBLICATIONS . THESE PUBLICATIONS
OFTEN INCLUDE INFORMATION WHICH IS NOT REPORTED IN THE
TRADITIONAL MASS MEDIA EFFORT . A LIST OF THE TRADE PUBLICA-
TIONS OF INTEREST TO THE CHEMICAL INDUSTRY IS INCLUDED .

42. PLEA FOR THE EXPANSION OF MEDICAL INFORMATION . THE EVER
INCREASING NEED FOR A MORE EFFICIENT ORGANIZATION OF MEDICAL
INFORMATION IN THE SOVIETx UNIONx INDICATES THAT THE PRESENT
DECENTRALIZED AND INEFFICIENT SYSTEM OF MEDICAL DOCUMENTATION
SHOULD BE REPLACED WITH A CENTRALIZED SYSTEM . THE CENTRALIZA-
TION OF MEDICAL DOCUMENTATION WOULD PERMIT A BETTER UTILIZATION
OF MEANS AND PERSONNEL ALREADY AVAILABLE TO VARIOUS ORGANIZA-
TIONS ENGAGED IN MEDICAL DOCUMENTATION AND BETTER DISSEMINATION
OF INFORMATION THROUGH IMPROVEMENT IN METHODS OF PRESENTATION
AND ENLARGEMENT OF THE SCOPE OF MEDICAL DOCUMENTATION .

43. GROWTH OF CHEMICAL LITERATURE : PAST , PRESENT , AND FUTURE .
ACTIVITY OF CHEMICALx ABSTRACTx SHOWS MARKED GROWTH :
JOURNALS COVERED BY CAx IN x1907 WERE x475 , IN x1960 ,
x9800 JOURNALS COVERED . NEW JOURNALS IN CHEMISTRY INCREASING
ABOUT x500 PER YEAR . IN x1960 THE x104,484 ARTICLES ABSTRACT-
ED CAME FROM x97 COUNTRIES IN x52 LANGUAGES . x27,675 PATENTS
WERE ABSTRACTED THE SAME YEAR . RUSSIANx CHEMICAL RESEARCH
PROGRESS IS SO GREAT THAT IF U.S.x AND U.S.S.R.x RATES OF
GROWTH CONTINUE UNCHANGED , THE U.S.S.R.x WILL OVERTAKE
THE U.S.x IN x1965 . SUPPORT IS NEEDED FOR DEVELOPING EXISTING
TOOLS AND DEVELOPING NEW INFORMATION SERVICES .

44. SOVIETx SCIENCE INFORMATION SERVICES . THIS IS A REPORT OF
U.S.x ABSTRACTING AND INDEXING SPECIALISTS WHO OBSERVED
THE OPERATIONS OF THE CENTRALIZED SCIENCE INFORMATION SERVIC
OF THE ALL-UNIONx INSTITUTEx OF SCIENTIFICx AND TECHNICALx
INFORMATIONx OF THE U.S.S.R.x . VINITIx IS PURPORTED TO BE THE
x1ST COMPLETELY CENTRALIZED INFORMATION SERVICE IN THE WORLD .

45. ON THE VARYING TASKS OF TECHNICAL LIBRARIANSHIP AND DOCUMENTA-
TION . A DESCRIPTION OF HUNGARYx 'S NETWORK OF TECHNICAL
LIBRARIES WITH THE HUNGARIANx CENTRALx TECHNICALx LIBRARYx
AS ITS CORE . AN ACCOUNT IS GIVEN OF TASKS CURRENTLY BEING
UNDERTAKEN , SUCH AS DECENTRALIZATION IN THE ACQUISITION OF
TECHNICAL LITERATURE , AND NATIONAL CONTROL OF SUBSCRIPTIONS
TO TECHNICAL PERIODICALS IN ORDER TO BUILD GREATER VARIETY
IN COLLECTIONS .

46. HISTORY AND FURTHER DEVELOPMENT OF SCIENTIFIC-TECHNICAL
INFORMATION , ESPECIALLY IN CZECHOSLOVAKIAx . THE RELATION-
SHIP OF THE LIBRARY TO SOCIETY IS BRIEFLY TRACED AND THE
DEVELOPMENT OF TECHNICAL LIBRARIES IS DISCUSSED . THE DEVELOP-
MENT OF SCIENTIFIC-TECHNICAL INFORMATION CENTERS IN CZECHO-
SLOVAKIAx AFTER x1945 IS DESCRIBED . THE MAIN FUNCTIONS OF
THESE CENTERS ARE OUTLINED . THE NUCLEUS CONSISTS OF THE
CENTRALx TECHNICALx LIBRARYx IN PRAGUEx , WHICH IS AUGMENTED
BY A TECHNICALx BRANCHx AND THE LIBRARYx OF THE BUREAUx
FOR STANDARDIZATIONx AND PATENTSx . THE DOCUMENTATIONx BRANCHx
OF THE MINISTRYx FOR PRODUCTIONx AND THE DOCUMENTATION BRANCHES
OF THE RESEARCH INSTITUTES OR OF THE LARGE PRODUCTION CENTERS
AS WELL AS THE TECHNICAL INDUSTRIAL LIBRARIES CONSTITUTE OTHER
CENTERS .

47. DOCUMENTATION AND ECONOMY . FUNDAMENTAL PROBLEMS OF DOCUMENTA-
TION ARE DISCUSSED FROM THE PROFESSIONAL AND ORGANIZATIONAL
POINT OF VIEW . IT IS STATED EMPHATICALLY THAT THE EXPERT MUST
BE PROVIDED WITH THE MATERIAL FOR HIS WORK WHATEVER THE FINANCIAL
COST . DEVELOPMENT IN DOCUMENTATION IN WESTx GERMANYx MUST
TAKE THE FORM OF COOPERATION BETWEEN INSTITUTIONS STAFFED
BY HIGHLY QUALIFIED PERSONNEL . COMPLETE CENTRALIZATION AS IN
THE SOVIETx UNIONx IS NEITHER POSSIBLE NOR DESIRABLE .
DOCUMENTARY ACTIVITY WILL BE COORDINATED AT NATIONAL LEVEL
BY THE PROPOSED INSTITUTx FUR DOKUMENTATIONSWESENx .

48. BUREAUx OF SHIPSx RAPIDx SELECTORx . BRIEFLY DESCRIBES THE
RAPIDx SELECTORx PROGRAM OF THE BUREAUx OF SHIPSx WHICH USES
A PROTOTYPE RAPIDx SELECTORx , DESIGNED TO SATISFY THE NEED
FOR INFORMATION RETRIEVAL ON A WHOLE-PAGE OR TOTAL-DOCUMENT
BASIS . THIS PROGRAM COMPLEMENTS THE BUREAU 'S WELL-ESTABLISHED
CARDOTYPEx INFORMATION RETRIEVAL PROGRAM , IN WHICH THE
INDIVIDUAL LINES AND PARAGRAPHS ARE THE RETRIEVAL UNITS
EMPLOYED MOST FREQUENTLY .

49. MAKING CLASSIFICATION SYSTEMS FOR PUNCHED-CARD CODING .
COUNTLESS EFFORTS HAVE BEEN MADE TO INVENT NEW CLASSIFICATION
SYSTEMS AND IMPROVE ON OLD CNES . THESE EFFORTS ARE BEING
INTENSIFIED TODAY BECAUSE OF THEIR APPLICATIONS IN MECHANIZED
SYSTEMS . SOME OF THE REQUIREMENTS AND ADVANTAGES OF CLASSIFICA-
TION SYSTEMS ARE DISCUSSED .

50. GUIDE TO MICROREPRODUCTION EQUIPMENT . PROVIDES TIMELY
AND ACCURATE FACTUAL COMPARABLE DATA , INCLUDING ILLUSTRA-
TIONS AND PRICES , FOR ALL KNOWN MICROREPRODUCTION EQUIPMENT
MANUFACTURED OR SOLD IN THE U.S.x . THIS x2ND EDITION UPDATES
THE INFORMATION FOUND IN ITS PREDECESSOR , ELIMINATES OBSOLETE
AND DISCONTINUED ITEMS AND EXTENDS COVERAGE TO MANY NEW AREAS
SUCH AS HIGH CAPACITY STORAGE AND RETRIEVAL AND DATA MANAGEMENT .
IT IS CLASSIFIED AND INDEXED .

APPENDIX 4.

THE FIRST FIVE ABSTRACTS OF JANASKE 1962 : UNEDITED VERSION.

1 ABE, K.

> The Japan Information Center of
> Science and Technology and its
> activities.
> Aslib Proc. 11(11):301-304. Nov. 1959.

A general description of the Japan Information
Center of Science and Technology which was es-
tablished for the purpose of contributing to the
development of science and technology by collect-
ing, processing, storing, and retrieving scientific
information. At the time of this report, the Center
was planning to change over to a mechanized storage
and retrieval system similar to that used by the U.S.
Patent Office. Services available from the Center
and the charges for each are listed.

2 ACKOFF, R.L.

> Report on a study of scientific com-
> munications for the National Science
> Foundation.
> In Shera, J.H., Kent, A., and
> Perry, J.W., ed. Information sys-
> tems in documentation, p. 66-74.
> New York, Interscience, 1957.

This reviews the initial reasoning and thinking be-
hind the design and construction of a project sup-
ported by the National Science Foundation, carried
out by the Operations Research Group at the Case
Institute of Technology to determine ways "to af-
fect the behavior of scientific institutions that are
responsible for the dissemination of recorded infor-
mation in order to achieve a net improvement in
scientific productivity."

3 ADAMS, S.

> Problems in communicating Russian
> science.
> Fed. Proc. 16(3):716-720. Sept. 1957.

An outline of the National Institutes of Health
Russian scientific translation program, with a dis-
cussion of the problems of selection, form, distri-
bution, acceptability, and translation rights. (:)ADI

4   ADDINALL, C.R., and STECHER, P.G.

Searching medicinal chemical litera-
ture.
Amer. Chen. Soc. Advn. Chem. Ser.
30:121-126. Ref. 1961.

Three representative problems of a manufacturer of
fine chemicals in searching for information in the
medical and chemical literature:searching the
botanical drug literature for folk remedies; inter-
preting and determining the content of foreign
prescriptions; and tracing and identifying the
Merck references.

5   ADKINSON, B.W.

The federal government and U.S.
scientific information.
J.Chem.Doc. 2(1):48-50. Jan. 1962.

In addition to presenting a general review of the
scope of the Federal Government's role in research
and development, and a review of the character
of the Federal Government's scientific information
activities, this article presents a review of the six-
point program which has been developed by the
National Science Foundation for improving the
dissemination of scientific information, particularly
among U.S. scientists and engineers. The author
concludes that the dessemination of scientific infor-
mation must be recognized as an integral part of
research and development.

BIBLIOGRAPHY

Allard, M., Elzière,M., Gardin, J.C., Hours, F.

  1963 Analyse conceptuelle du Coran sur cartes perforées.
       Paris, La Haye, Mouton, t. I, Code; t. II, Commentaire.

Alouche, F., Bély,N., Gardin, J.C., Lévy, F., Perriault, J.

  1967 Economie générale d'une chaîne documentaire mécanisée.
       Paris, Gauthier-Villars.

Annamalai, E.

  1970 "On Moving from a Coordinate Structure in Tamil".
       In Papers from the Sixth Regional Meeting Chicago
       Linguistic Society, Chicago Linguistic Society.
       Chicago, Illinois, 131-146.

Annear Thompson, Sandra.

  1970a "Relative Clause Structures and Constraints on
        Types of Complex Sentences". In Working Papers in
        Linguistics N° 6, The Ohio State University,
        Columbus, Ohio, 20-40.

  1970b "The Deep Structure of Relative Clauses". In
        Working Papers in Linguistics N° 6, The Ohio
        State University, Columbus, Ohio, 41-58.

Austin, J.L.

  1962 How to Do Things with Words. New York, Oxford
       University Press.

Bach, Emmon.

  1965 "Linguistique structurelle et philosophie des sciences".
       Diogène, 51, 117-136.

  1968 "Nouns and Noun Phrases". In E.Bach and R. Harms, eds.,
       Universals in Linguistic Theory. New York, Holt,
       Rinehart and Winston, 90-122.

Bach, Emmon and Harms, Robert, eds.

1968 <u>Universals in Linguistic Theory</u>. New York, Holt,
Rinehart and Winston.

Ballard, D. Lee. Conrad, R.J., Longacre, R.E.

1971 "The Deep and Surface Grammar of Interclausal
Relations". <u>Foundations of Language</u>, 7, 1, 70-118.

Bély, N., Borillo, A., Slot-Decauville, N., Virbel, J.

1970 <u>Procédures d'analyse sémantique appliquées à la
documentation scientifique</u>, Paris, Gauthier-Villars.

Berelson, B.

1952 <u>Content Analysis in Communication Research</u>. Glencoe,
Ill., Free Press (quoted by Stone <u>et al.</u> 1966).

Bierwisch, Manfred and Kiefer, Ferenc.

1969 "Remarks on Definitions in Natural Language". In
F. Kiefer, ed., <u>Studies in Syntax and Semantics</u>.
Dordrecht, Holland, Reidel, 55-79.

Binkert, Peter J.

1970 <u>Case and Prepositional Constructions in a
Transformational Grammar of Latin</u>. Unpublished ms.

Bloomfield, L.

1933 <u>Language</u>. New York, Holt, Rinehart and Winston.

Bolinger, Dwight.

1965 "The Atomization of Meaning". <u>Language</u>, 41, 555-573.

Botha, R.P.

1968 <u>The Function of the Lexicon in Transformational
Generative Grammar</u>. The Hague-Paris, Mouton.

Braithwaite, R.B.

1962 "Models in the Empirical Sciences". In E. Nagel, P.
Suppes, and A. Tarski, eds., <u>Logic, Methodology and
Philosophy of Science</u>. Stanford, California. (quoted
by Botha 1968).

Chauvineau, Jean.

>1962 *La logique moderne*. Paris, Presses Universitaires de
France.

Cherry, Colin.

>1957 *On Human Communication, A Review, a Survey, and a
Criticism*. 1961 reprint, New York, Science Editions.

Chomsky, Noam.

>1964 *Current Issues in Linguistic Theory*. The Hague,
Mouton.

>1965 *Aspects of the Theory of Syntax*. Cambridge, Mass.,
M.I.T. Press.

>1967 "Remarks on Nominalization". Unpublished ms.
Reprinted in R.A. Jacobs and P.S. Rosenbaum,
Readings in English Transformational Grammar,
Ginn and Company, 1970, 184-221.

>1969a "Form and Meaning in Natural Language". In J.D.
Roslansky, ed., *Communication*. Amsterdam, North-
Holland, 65-85.

>1969b "Deep Structure, Surface Structure, and Semantic
Representation". Reproduced by the Indiana Univer-
sity Linguistics Club.

Coyaud, Maurice.

>1965 "Document Automatic Indexing with the help of
Semantic Information". Paper presented at the
International Conference on Computational Linguistics.
New York, 1965.

>1966 *Introduction à l'étude des langages documentaires*.
Paris, Klincksieck.

Coyaud, M., Siot-Decauville, N.

>1967 *L'analyse automatique des documents*. Paris, La Haye,
Mouton.

Cros, R.C., Gardin, J.C., Lévy, F.

    1964 L'automatisation des recherches documentaires, un
        modèle général : le SYNTOL. Paris, Gauthier-Villars.
        (2nd ed., 1968).

Dougherty, Ray C.

    1970 "Recent Studies on Language Universals". Foundations
        of Language, 6, 4, 505-561.

Dubois, Jean.

    1963 Le vocabulaire politique et social en France de
        1869 à 1872. Paris, Larousse (reviewed by Mounin
        1965).

Elliot, D.E., Huang, Sh., Langendoen, D.T., Lee, P.G., and
Lehiste, Ilse.

    1969 Working Papers in Linguistics N° 3. The Ohio State
        University, Columbus, Ohio.

Fillmore, Charles J.

    1965 "Entailment Rules in a Semantic Theory". Project
        on Linguistic Analysis, RF 1685-3 Report N° 10,
        The Ohio State University, Columbus, Ohio.

    1967 "The Grammar of HITTING and BREAKING". In Working
        Papers in Linguistics N°1, 9-29. Reprinted in R.A.
        Jacobs and P.S. Rosenbaum, eds. Readings in English
        Transformational Grammar. Waltham, Mass., Ginn and
        Company, 1970, 120-133.

    1968a "The Case for Case". In E. Bach and R. Harms, eds.,
        Universals in Linguistic Theory. New York, Holt,
        Rinehart and Winston, 1-88.

    1968b "Lexical Entries for Verbs". In Working Papers in
        Linguistics N° 2, The Ohio State University,
        Columbus, Ohio, 2-29. Also in Foundations of
        Language, 4, 4, 373-93.

    1968c "Review of "Componential Analysis of General Vocab-
        bulary : the Semantic Structure of a Set of Verbs
        in English, Hindi and Japanese", by Bendix. (I.J.A.L.,
        Vol. 32, N° 2, Publication 41, 1966)". In Working
        Papers in Linguistics N° 2, The Ohio State Univer-
        sity, Columbus, Ohio, 30-64.

1969a "Types of Lexical Information". In F. Kiefer, ed., **Studies in Syntax and Semantics**, Dordrecht, Reidel, 109-132.

1969b "Verbs of Judging : an Exercise in Semantic Description". **Papers in Linguistics**, 1, 91-115.

1970a "Verbes de jugement. Essai de description sémantique". **Langages**, 17, 56-72.

1970b **Course on Syntax**. Linguistic Institute of the Linguistic Society of America, Columbus, Ohio, unpublished ms.

1970c "Required Reading for Linguistics 820". Linguistic Institute of the Linguistic Society of America, Columbus, Ohio, unpublished ms.

Fillmore , C.J. and Lehiste, Ilse, eds.

1968 **Working Papers in Linguistics N° 2**, The Ohio State University, Columbus, Ohio.

Firth, J.R.

1964 **The Tongues of Men and Speech**. London, Oxford University Press (see Langendoen 1968).

Gardin, J.C.

1965 "SYNTOL". In **Rutgers Series on Systems for the Intellectual Organization of Information**, II. New Brunswick, N.J.

1966 "L'utilisation de l'information sémantique en documentation automatique". In **Premier Colloque international de linguistique appliquée**. Nancy, Faculté des Lettres et des Sciences Humaines.

1969 "Semantic Analysis Procedures in the Science of Man". **Social Science Information**, VIII-1, The Hague, Mouton.

Gardin, Natacha.

1969 "Le lexique intermédiaire. Un nouveau pas vers la coopération internationale dans le domaine de l'information scientifique et technique". **Bulletin de l'Unesco à l'intention des bibliothèques**, XXIII, 2, 66-71.

Gardin, N., Lévy, F.

1967 Etude de concordance interlexicale dans le domaine
de l'information scientifique et technique. Marseille,
Groupe d'Etude sur l'Information Scientifique,
unpublished ms.

Gochet, Paul.

1970a (MS of) Logique, langage et universaux : Esquisse
d'une théorie nominaliste de la proposition. Paris,
Armand Colin (Forthcoming).

1970b Notions de logique. Université de Liège.

Goldin, Mark G.

1968(?) Spanish Case and Function. Washington, Georgetown
University Press.

Green, Georgia M.

1968 "On TOO and EITHER, and not just on TOO and EITHER,
either". In Bill J. Darden, C.-J.N. Bailey and Alice
Davison, eds., Papers from the 4th Regional Meeting,
Chicago Linguistic Society, Chicago Linguistic
Society, University of Chicago, Chicago, Illinois,
22-39.

Harris, Zellig S.

1965 String Analysis of Sentence Structure. The Hague,
Mouton (Third printing).

1968 Mathematical Structures of Language. New York,
Wiley-Interscience.

1970 "Linguistic Transformations for Information Retrieval".
In Papers in Structural and Transformational Lin-
guistics, Dordrecht, Reidel, 458-471. First published
in Proceedings of the International Conference on
Scientific Information (1958), 2, NAS-NRC, Washington
D.C. 1959.

Hempel, C.G.

    1952 <u>Fundamentals of Concept Formation in Empirical
    Science</u>. International Encyclopedia of Unified
    Science, Vol. II, N° 7, Chicago (quoted in Botha
    1968).

Henderson, M.M., Moats, J.S., Stevens, M.E., Newman, S.M.

    1966 <u>Cooperation, Convertibility, and Compatibility
    among Information Systems : a Literature Review</u>.
    Washington, National Bureau of Standards. (Miscella-
    neous Publication 276.) (quoted in J.C. Gardin 1969)

Hofman, Th. R.

    1969 "Underlying vs Superficial Grammatical Relations".
    <u>The Canadian Journal of Linguistics</u>, 15:1, 11-26.

Hornby, A.S.

    1954 <u>A Guide to Patterns and Usage in English</u>. London,
    Oxford University Press (Twelfth impression 1970).

    1963 <u>The Advanced Learner's Dictionary of Current
    English</u>. London, Oxford University Press.

Janaske, Paul C., ed.

    1962 <u>Information Handling and Science Information.
    A Selected Bibliography 1957-1961</u>. Washington,
    D.C., American Institute of Biological Sciences.

Jespersen, Otto.

    1965 <u>The Philosophy of Grammar</u>. New York, Norton and
    Company (first published in 1924).

    1966 <u>Essentials of English Grammar</u>. London, George
    Allen and Unwin (first published in 1933).

Katz, J.J.

    1967 "Recent Issues in Semantic Theory". <u>Foundations
    of Language</u>, 3, 124-194.

Katz, J.J. and Fodor, J.A.
1963 "The Structure of a Semantic Theory". Language, 39, 170-210.

Katz, J.J., Fodor, J.A.
1964 " The Structure of a Semantic Theory". In Fodor and Katz, eds., The Structure of Language, Readings in the Philosophy of Language. Englewood Cliffs, N.J., Prentice Hall, 479-518.

Katz, J.J. and Postal, P.M.
1964 An Integrated Theory of Linguistic Descriptions. Cambridge, Mass., M.I.T. Press.

Kuno, Susumu.
1966 "The Predictive Analyzer". In David G. Hays, ed., Readings in Automatic Language Processing. New York, American Elsevier, 83-106.

Kuroda, S.Y.
1968 "English Relativization and Certain Related Problems". In D.A. Reibel and S.A. Schane, Modern Studies in English Readings in Transformational Grammar. Englewood Cliffs, N.J., Prentice-Hall, 1969, 264-287.

Lakoff, George.
1968a "On the Nature of Syntactic Irregularity". In Linguistic Institute Packet of Papers, 1968 Linguistic Institute of the Linguistic Society of America, University of Illinois. (originally published by The Computation Laboratory of Harvard University Mathematical Linguistics and Automatic Translation, Report No. NSF-16, Cambridge, Mass., 1965).

1968b "Instrumental Adverbs and the Concept of Deep Structure". Foundations of Language, 4, 4-29.

1969a "On Derivational Constraints". In R.I. Binnick,
A. Davison, O.M. Green and J.L. Morgan, eds.,
<u>Papers from the Fifth Regional Meeting of the
Chicago Linguistic Society</u>, Chicago Linguistic
Society, Chicago, Illinois, 117-139.

1969b "Presuppositions and Relative Grammaticality".
In Todd, Williams, ed., <u>Philosophical Linguistics.</u>
Series I, Great Expectations (Evanston), 103-116.

1969c "On Generative Semantics". Unpublished ms. To
appear in D.D. Steinberg and L.A. Jakobovits, eds.,
<u>Semantics - An Interdisciplinary Reader in Philo-
sophy, Linguistics, Anthropology, and Psychology.</u>
London, Cambridge University Press.

1970 "Natural Logic and Lexical Decomposition". In
<u>Papers from the 6th Regional Meeting, Chicago
Linguistic Society</u>, Chicago, Illinois, 340-362.

1971 Introduction to the version of <u>Linguistics and
Natural Logic</u> to appear in J. Lehler (ed.),
<u>Handbook of Cognitive Psychology</u>. University of
Michigan, Ann Arbor, Michigan, unpublished ms.

Lakoff, George and Peters, Stanley.

1969 "Phrasal Conjunction and ay metric Predicates".
In D.A. Reibel and S.A. Schane, <u>Modern Studies in
English. Readings in Transformational Grammar.</u>
Englewood Cliffs, N.J., Prentice-Hall, 113-142.

Lakoff, Robin.

1970 "IF's, AND's, and BUT's about Conjunction". Revised
version of a paper read at the Ohio State University
Spring Semantics Festival, April 1969, unpublished
ms.

Langacker, Ronald W.

1967 <u>Language and its Structure</u>. New York, Harcourt,
Brace and World.

Langendoen, D. Terence.

> 1967 "On Selection, Projection, Meaning, and Semantic Content". In <u>Working Papers in Linguistics N° 1</u>. The Ohio State University, Columbus, Ohio, 100-108.

> 1968 <u>The London School of Linguistics : A Study of the Linguistic Theories of B. Malinowski and J.R. Firth</u>. Cambridge, Mass., The M.I.T. Press.

> 1969 <u>The Study of Syntax. The Generative-Transformational Approach to the Structure of American English</u>. New York, Holt, Rinehart and Winston.

> 1970 <u>Essentials of English Grammar</u>. New York, Holt, Rinehart and Winston.

Lasswell, H.D., Lerner D. and de Sola Pool, I.

> 1952 <u>The Comparative Study of Symbols</u>. Stanford, Calif., Stanford University Press (quoted in Stone <u>et al.</u> 1966).

Lawlor, Reed C.

> M.S. <u>Training Manual</u>. University of Southern California, California, unpublished ms.

Lee, Gregory.

> 1971 "Subjects and Agents : II". In <u>Working Papers in Linguistics N° 7</u>, Computer and Information Science Research Center, The Ohio State University, Columbus, Ohio, L1-L118.

Leech, G.N.

> 1966 <u>English in Advertising. A Linguistic Study of Advertising in Great Britain</u>. London, Longmans.

> 1968 "Some Assumptions in the Metatheory of Linguistics". <u>Linguistics</u>, 39, 87-102.

> 1969 <u>Towards a Semantic Description of English</u>. London, Longmans.

Levin, Samuel R.

    1971 "The Analysis of Compression in Poetry". Foundations
         of Language, 7, 1, 38-55.

Linsky, Leonard.

    1967 Referring. New York, Humanities Press.

Lyons, John.

    1963 Structural Semantics. An Analysis of Part of the
         Vocabulary of Plato. Oxford, Basil Blackwell.

    1968 Introduction to Theoretical Linguistics. Cambridge,
         G.B., Cambridge University Press.

McCawley, James.

    1968 Review of Sebeok (ed.) "Current Trends in Lin-
         guistics, III : Theoretical Foundations". Language,
         44, 3, 556-593.

    1970 "English as a VSO Language". Language, 46, 2, 286-
         299.

Moravcsik, J.M.E.

    1967 "Linguistic theory and the Philosophy of Language".
         Foundations of Language, 3, 209-233.

Mounin, Georges.

    1965 Review of J. Dubois 1963. La linguistique, 1,
         131-137.

Nagel, Ernest.

    1961 The Structure of Science. Problems in the Logic
         and Scientific Explanation. London (quoted by
         Botha 1968).

Noël, Jacques.

    1968 Linguistic Problems in Mechanized Indexing of
         English Abstracts. University of Victoria, Victoria,
         B.C., unpublished M.A. thesis.

    1971 "La notion de compétence linguistique : de
         Chomsky à quelques travaux récents". Part I, ITL,
         11, 31-49 (Part II, forthcoming).

Piaget, Jean.

    1968 <u>Le structuralisme</u>. Paris, Presses Universitaires de France.

Popper, K.R.

    1963 <u>Conjectures and Refutations. The Growth of Scientific Knowledge</u>. London (quoted by Botha 1968).

Postal, Paul.

    1967 "Crazy Notes on Restrictive Relatives and Other Matters" York:o:m,Heights,N.Y.,Thomas J. Watson Research Center, IBM (quoted in Annamalai 1970).

    1969 "Anaphoric Islands". In <u>Papers from the Fifth Regional Meeting Chicago Linguistic Society</u>, Chicago Linguistic Society, Chicago, Illinois, 205-239.

Pottier, Bernard.

    1968 "Champ sémantique, champ d'expérience et structure lexicale". <u>Zeitschrift für französische Sprache und Literatur</u>. Beiheft, Neue Folge, Heft 1, herausgegeben von W. Theodor Elwert, <u>Probleme der Semantik</u>. Wiesbaden, Franz Steiner Verlag, 37-40.

Reichenbach, H.

    1947 <u>Elements of Symbolic Logic</u>. New York, The Free Press.

Reichling, A.

    1965 <u>Verzamelde studies over hedendaagse problemen der taalwetenschap</u>. Derde herziene en uitgebreide druk. Zwolle (quoted in Botha 1968).

Rivetti-Barbo, F.

    1965 "Sense, Denotation, Context of Sentences". In A.T. Tymencka, ed., <u>Contributions to Logic and Methodology in Honor of J.M. Bochenski</u>. Amsterdam, North-Holland, 208-242 (quoted in Gochet 1970a).

Robinson, R.

    1950 <u>Definition</u>. Oxford, Clarendon Press (quoted in Bierwisch and Kiefer 1969).

Ross, John R.

    1967 "Constraints on Variables in Syntax". M.I.T.,
        unpublished doctoral dissertation. In <u>Linguistic</u>
        <u>Institute Packet of Papers</u>, 1968 Linguistic
        Institute of the Linguistic Society of America,
        University of Illinois.

    1970 "On Declarative Sentences". In R.A. Jacobs and
        P.S. Rosenbaum, <u>Readings in English Transformational</u>
        <u>Grammar</u>, Ginn and Company, Waltham, Mass., 222-272.

Sager, Naomi.

    1967 "Syntactic Analysis of Natural Language". In F. Alt
        and M. Rubinoff, eds., <u>Advances in Computers 8</u>,
        New York, Academic Press, 153-188.

Salkof, M. and Sager, N.

    1967 "The Elimination of GrammaticalRestrictions in a
        String Grammar of English". Paper presented at the
        <u>2e Conférence Internationale sur le Traitement</u>
        <u>Automatique des Langues</u>, Grenoble.

Salton, Gerard and Lesk, M.E.

    1968 "Computer Evaluation of Indexing and Text Processing".
        <u>Journal of the Association for Computing Machinery</u>,
        15.1, 8-36.

Salton, Gerard.

    1966 "Automatic phrase matching ". In David G. Hays, ed.,
        <u>Readings in Automatic Language Processing</u>. New
        York, American Elsevier, 169-188.

    1968 <u>Automatic Content Analysis in Information Retrieval</u>.
        Technical Report No. 68-5. Department of Computer
        Science, Cornell University, Ithaca, N.Y.

Saussure, F. de

    1968 <u>Cours de linguistique générale</u>. Paris, Payot.

Shapiro, P.A.

    1967 "ACORN - An Automated Coder of Report Narrative".
        Methods of Information in Medicine, 6.4, 153-162.

Simmons, Robert F.

    1970 "Natural Language Question-Answering Systems :
        1969". Communications of the ACM, 13, 1, 15-30.

Stockwell, Robert, Schachter, Paul, Partee, Barbara.

    1968 Integration of Transformational Theories on
        English Syntax. US Air Force, ESD-TR-68-419.

Stone, P.J., Dumphy, D.C., Smith, M.S., Ogilvie, D.M.

    1966 The General Inquirer : a Computer Approach to
        Content Analysis. Cambridge, Mass., MIT Press.

Strawson, P.F.

    1950 "Phrase et acte de parole". (Translated from English)
        Langages, 17, mars 1970, 19-33.

Vinay, J.P. et Darbelnet, J.

    1958 Stylistique comparée du français et de l'anglais.
        Méthode de traduction. Paris, Didier.

Weinreich, U.

    1966 "Explorations in Semantic Theory". In T.A. Sebeok,
        ed., Current Trends in Linguistics, III. The Hague,
        Mouton, 394-477.

West, Michael.

    1965 An International Reader's Dictionary. London,
        Longmans.

Zandvoort, R.W.

    1966 A Handbook of English Grammar. London, Longmans.